PERFORMANCE MEASURES FOR HUMAN FACTORS ENGINEERING EVALUATION OF EARL EQUIPMENT

FINAL REPORT

dai

PREPARED FOR:

EDGEWOOD ARSENAL RESEARCH LABORATORIES

EDGEWOOD, MARYLAND

CONTRACT NO. DA18-035-AMC-702 (A)

Distribution Statement

Distribution of this document is unlimited.

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Disposition

Destroy this report when no longer needed. Do not return it to the originator.

#34

SSD66-332 (618)

Final Comprehensive Report

PERFORMANCE MEASURES FOR HUMAN FACTORS ENGINEERING EVALUATION OF EARL EQUIPMENT

Prepared for

Edgewood Arsenal Research Laboratories
Edgewood, Maryland

Contract No. DA18-035-AMC-702 (A)

30 June 1966

Prepared by

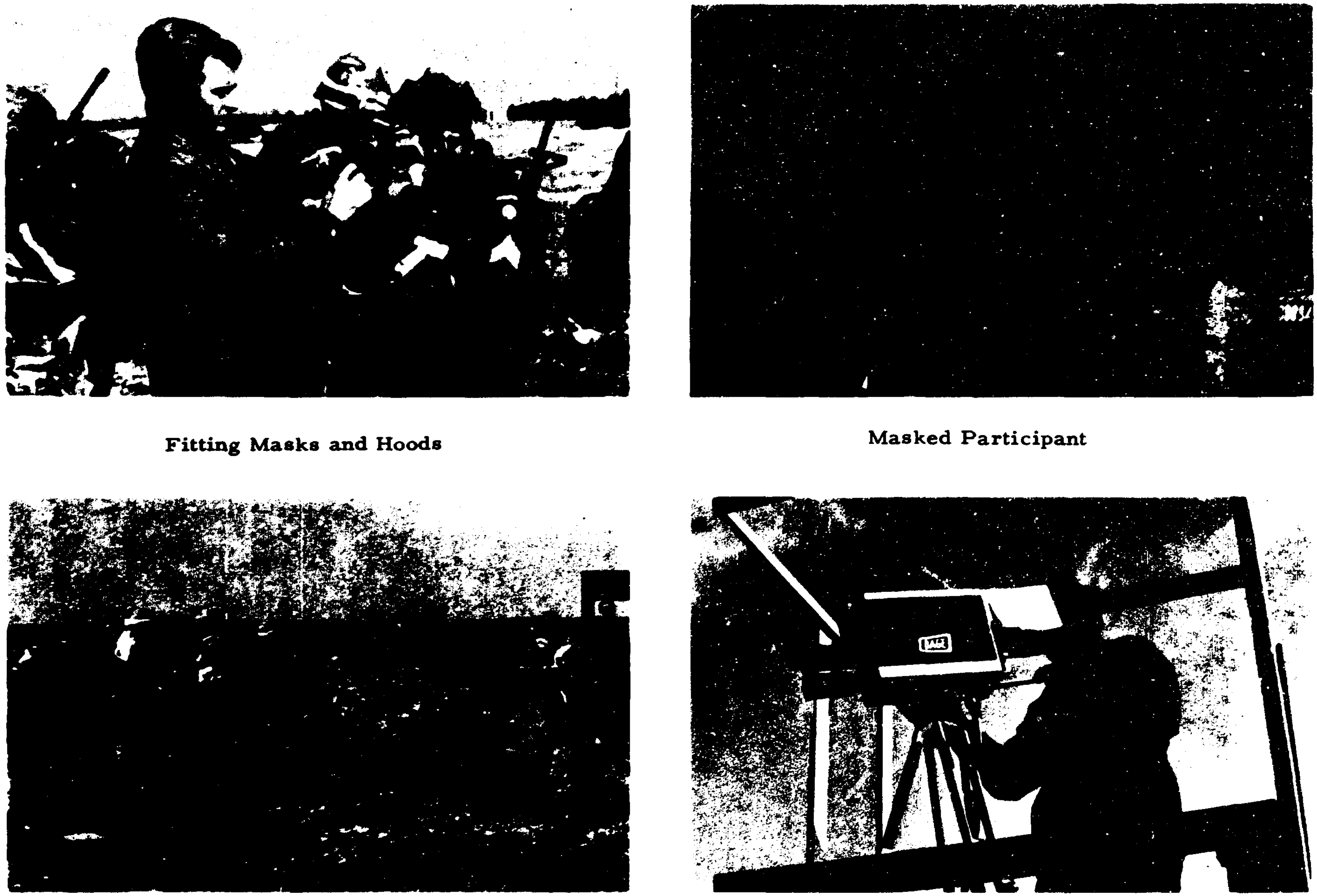
Dunlap and Associates, Inc.
One Parkland Drive
Darien, Connecticut 06821

The Problem and the Objective

The tactical team, in modern warfare, performs an increasing proportion of military missions in relation to the individual soldier. Individual field performance has been the object of test design for longer and more often than the interactivities of the tactical team.

Respiratory protective equipment has been exhaustively field-tested for physiological effects on individuals, but only rarely for its overall effect on a tactical team in accomplishing its mission. No standard team test of protective gear exists.

This project attempts to recommend standard team testing techniques that will generate useful, reliable data about the effects of EARL protective masks and hoods on small Army teams performing tactically.



Fitting Masks and Hoods

Masked Participant

Team On-Site

Videorecording of Activity

Figure 1

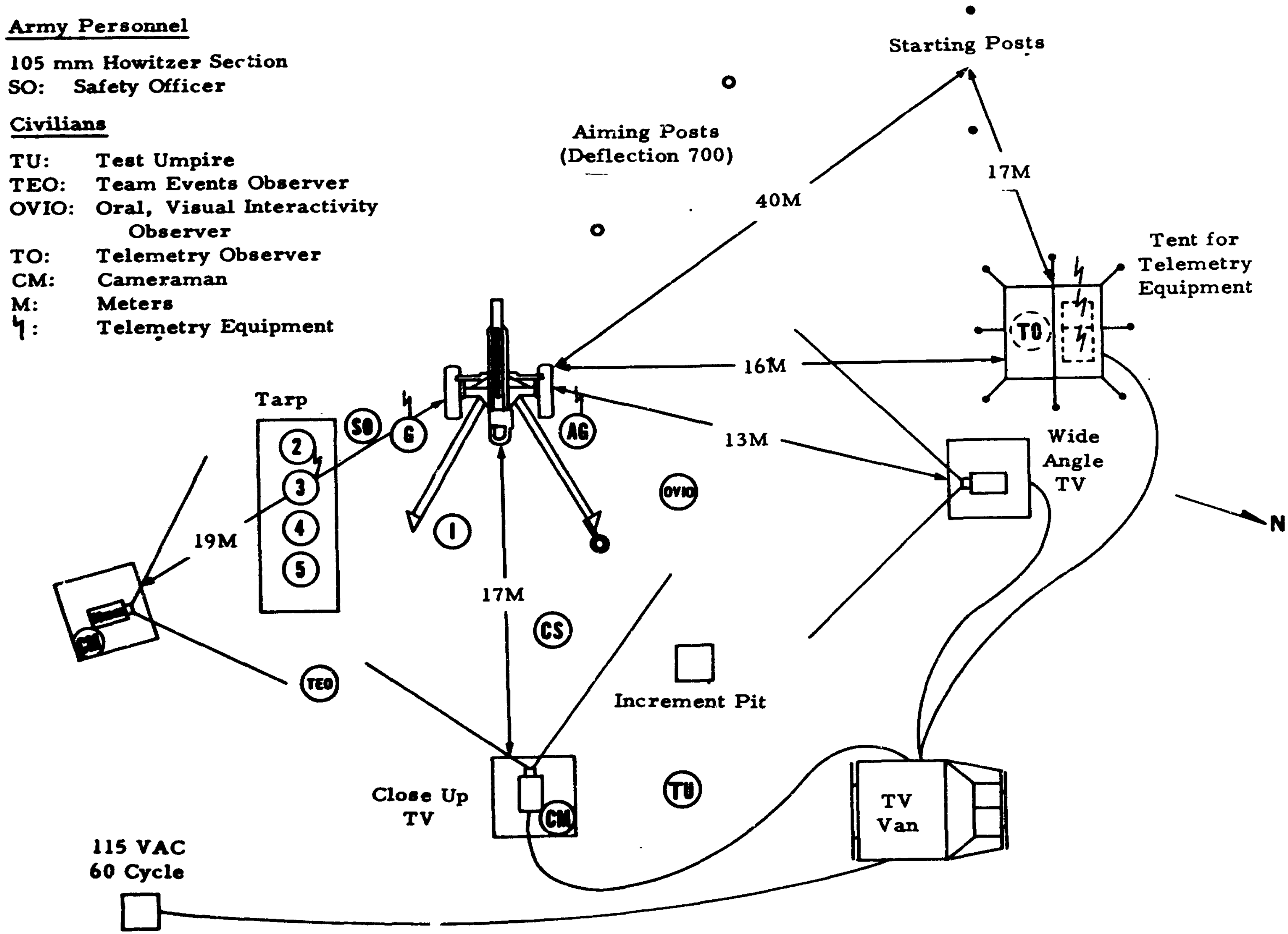


Figure 2. Field test firing and observation site.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

PROJECT PERSONNEL

Throughout the course of this study, the following staff was active in the design development and successful conclusion of the project:

Project Contract Officer for Edgewood Research Labs:

John W. Lewis, Human Engineering Research Branch
Samuel Jackson, Assistant

Contractor Staffing for Dunlap and Associates, Inc.:

Responsible Officer	- Roland C. Casperson, Ph. D.
Project Director	- Rod E. Packer, Ph. D.
Project Staff	- James Roden
	- John Hamilton
Technical Assistants	- Judy Sherwood
	- Bonnie Kaplan

At various times during the study, the following persons lent their support on critical points:

Edgewood Research Labs; Edgewood Arsenal, Md.

Major E. R. Clovis, Experimental Medical Labs
Lloyd Matter, Naval Medical TV Facility
Peter Allera & Sam McDowell, 16 mm Cameramen, Edgewood

Dunlap and Associates, Inc., Darien, Connecticut

Treadway Parker	- Associate Psychologist
Frank Affinito	- Telemetry Engineering
Robert McCay	- Senior Psychologist
Howard We[illegible]s	- Statistical Assistant
Jane Bowen	- Data Reduction Assistant
James Link	- Staff Scientist

I. T. V., Inc., New York City

Joe Bashlow, Television Recording Technician
Ivar Moorae, Television Cameraman

Fort Monroe, Virginia, CONARC

Col. Mollen and Lt. Col. Crispell, CBR Office

SECTION I

INTRODUCTION

A. Contractual Task

The Statement of Work for Contract No. DA18-035-AMC-703(A) describes the objective of this investigation as the development of, ". . . field performance measures to be used at the squad or similar level for human factors evaluation of items or equipment under development and test by U. S. Army Edgewood Arsenal Chemical Research and Development Laboratories."* To obtain this objective, five phases of study effort were prescribed in accordance with the technical approach proposed by Dunlap and Associates, Inc. in our study plan:

1. Select representative team task(s), in collaboration with the Contract Project Officer, for study.

2. Analyze task(s) on a time and error basis as a means of predicting performance degradation resulting from the use of protective equipment.

3. Test hypotheses concerning performance effects under field conditions.

4. Develop standard field performance measures, scoring techniques and methods for evaluating protective equipment.

5. Develop a sound film depicting standard task test format and observation techniques.

In brief, as the following sections of this report will clarify, each of these steps was undertaken and completed, and the primary objective of this study has been satisfied. The authors of this report feel, moreover, that an equally important product of this investigation, and one which should have great general application in the study of human performance in the field, is the technique developed for recording and analysis of the field data using video techniques. Sections IV, V, and VI of this report discuss more fully the relative advantages and disadvantages of the various techniques used in this and other studies for obtaining, processing and analyzing field data. At this point it is sufficient to say that the use of television cameras and video-recording make the task of obtaining meaningful and reliable field measures

---

* The previous designation for U. S. Army Edgewood Arsenal Research Laboratories prior to 1966 reorganization.

of performance of entire teams and of the individuals within them a feasible objective, rather than the unwieldy and even unrealistic task it has so often proved to be in the past.

B. Background

U. S. A. Edgewood Arsenal has the continuing responsibility to improve the design of chemical agent protective devices. Their success in this mission is clearly evident from comparison of the cumbersome, relatively inefficient gas masks provided our armed forces in World War I with the protective equipment [redacted] available for use today. In fact, the design of this equipment has improved to a point where even the direct advantages and disadvantages of a design modification are not always clearly apparent to those responsible for their evaluation. With the additional requirements for protection against widely divergent chemical, biological and radiational agents and the increasing complexity of the soldier's tactical activities in the field, earlier physiological and performance tests carried out with the individual soldier in the laboratory are inadequate as comprehensive evaluative techniques. These factors, combined with the awareness that team performance is not simply an additive function of the performance of those individuals who make up the team, stimulated the interest of the Human Factors Engineering Branch, Experimental Medicine Division, at Edgewood Arsenal to undertake this study.

As a result of Dunlap and Associates, Inc., varied experience in military evaluation studies, and in view of the stated problem, three significant problem areas critical and unique to our purpose determined the direction of our study efforts.

(1) Team performance measurement, generally, has not been dealt with to any great extent in the field situation. Although attempts have been made to measure and evaluate group performance in the controlled (laboratory) situation or in a "field" environment of a manufacturing plant, little has been done to quantify and assess the more dynamic interactions characteristic of a military team in the operational or actual field situation. For this reason, the development of an effective and efficient methodology for obtaining data concerning team field performance is of critical importance and, in itself, would provide sufficient justification for a study of this nature. Since field operating conditions are, by definition, impossible to control in a manner desirable for experimental study, techniques for obtaining the maximum amount of pertinent data in a reliable form must be exploited to the fullest. The techniques developed for this investigation have attempted to accomplish this goal.

(2) Task selection is critical to the validity of the results obtained. Although it is conceivable to construct test tasks having all of the

psychophysical or sensory-motor components of the real world, much is lost in the translation. Face validity, often demeaned as a critical item in testing, is vital to the test situation if meaningful results are to be obtained for application to combat. Whether or not the research personnel can accept and justify the use of "substitute" tasks, as measures of observable performance, is not the important issue. The emotional tone, the motivation, and the stresses of a "real" task cannot be simulated and the results, therefore, can always be questioned. For this reason, an actual field situation was selected as the vehicle for our investigation.

equipment

(3) Protective [redacted] has been developed to a point of sophistication where design improvement may not have discriminable effects on the normal operations of military personnel. For this reason, again, it would seem critical to employ a test situation as close to operating conditions as possible, assuring by careful data collection and interpretation that all meaningful performance measures are available for study.

The task chosen was a firing problem for a 105 mm howitzer crew.

C. Preparations

Several preparatory steps were taken prior to undertaking the study reported in the following pages.

Initially, a literature search was undertaken to familiarize the staff with the techniques used in and results of previous team performance studies conducted in the field. Although there have been many field exercises reported in the literature, the data collection and analyses reported for them provided little useful information for the study. Studies conducted in the field, involving the use of protective clothing, have been even less well controlled. There is general agreement, however, that masks interfere with oral communications. Background gained from the literature search will be discussed in detail in Section II.

Early in the contract period, a visit to Camp Pickett was scheduled to observe the methods being used in measuring the performance of individual infantrymen under simulated combat conditions. It was made quite clear from these observations that obtaining valid performance data in the field for single individuals is a problem--the difficulty of which is compounded when a team of individuals is the subject of concern.

We decided, in agreement with the Contract Project Officer at a meeting at Edgewood Arsenal, to use the activities of the howitzer gun crew as our model for team performance. To this end, detailed Operational Sequence

Diagrams were developed for the complete RSOP exercise including live firing of the 155 mm howitzer. Later, as a result of a trip to Fort Sill, Oklahoma, it was decided that the 105 mm howitzer gun crew would best satisfy our criteria and still permit us to meet the schedule established for our study plan.

SECTION II

INITIAL DESIGN

With the overall objectives and direction of the study effort well delineated, the development of a tentative design for the appropriate performance testing was begun at Dunlap and Associates.

The developmental design of a field evaluation technique for team performance with protective gear is a specific new area. However, the various aspects of the problem have received considerable attention previously. Accordingly, the general literature, in areas of test structuring, field observation, team performance, etc. was surveyed first to discover where developmental design already done could be applied. Then the specific design of both observer and participant activities was outlined with provision for an actual field test of this tentative structure.

Several particular topics of concern in this effort were singled out. One was the probability that communications among team members would be at least as important a variable as time and error in evaluating a group performing with masks over eyes and mouth. Another was the probability that the subtle pattern of team activity should rightly be analyzed in depth to avoid the loss of diagnostic capabilities that occurs in gross averaging of data from sub-tasks into the overall task and across interacting individuals. A third topic was the problem of objectivity in evaluating "quality" of performance, controversial enough in a tactical activity to suggest competitive approaches: "expert" ratings, time/error ratios, firing accuracy, etc. These topics of unique interest to the study, plus the broader area of general task design, team evaluation, field testing techniques and protective equipment evaluation were carefully investigated.

Following are descriptions, first, of the rather small core of pertinent published research encountered, and then, of the initial design as it was developed by the study staff and consultants prior to any modification and sophistication through actual field applications. During this design phase of the study, provision was simultaneously being made to field test the entire tentative design in order to move readily on to its realistic and knowledgeable modification into a practical technique.

A. <u>Findings from Research Survey</u>

Following are summaries of pertinent data found in the published literature. References cited may be found in full in the bibliography, Appendix A.

1. Field Testing

Field tests and studies are rapidly assuming a role of major importance within the experimental research area (Rabideau, 1964). The outstanding point of the field testing approach has been the expense involved in covering test exercises. Man time, preparation, liaison, and equipment, have been very costly (Rabideau, 1964). Field tests have been a problem in other ways. Relatively poor control over many variables such as weather, and malfunctions of system devices or recording devices have occurred in the field. Also, control of test subjects availability is usually in hands other than the experimentors, and problems of priorities may arise. The weak control over selection and use of test subjects may often cast serious doubt on the generalizability of the data obtained.

Field measurement of human performance has often been in relation to man-machine systems and generally restricted to systems analysis and the generation of reliability data for overall system effectiveness. This type of study involves the role of the human operator's contribution to system reliability. The accumulation of data over a number of trials then permits an effective prediction of system reliability based on the type of human behavior involved in the task. A broad interpretation of this approach is that any task involving a human operator and a machine (such as the howitzer) can be considered as a total system. The reliability of this system must include the human performer, individually or as a member of a team (Meister, 1964). The reliability of total mission performance must include the reliability of individual or team behaviors, and this involves the study of human potentiality for failure.

In the case related by Meister (1964) performance reliabilities for task elements are progressively combined through the use of a series product rule to yield a total reliability estimate, usually stated as probability estimate of error. This deals with man-machine interface, and develops relationships between human factors and reliability engineering as interrelated disciplines.

Another approach reported by Swain (1964) employs the use of a large background of reliability studies of more accurate estimates of error rates and compares these to similar equipment under study. The human factors specialist then relates tasks in the system at hand to similar tasks for which he has error rate data. This is the theoretical foundation of a data store or "data bank." The impetus for the data bank has been the great deficiency in available data of estimates for a large number of human behavioral elements (Meister, 1964).

Any study contributing to the data bank must be relatively large in magnitude. The field data accumulated in the Air Force Category II field tests is so considerable in bulk that it is required that automated

electronic data processing be used to assure reporting in a reasonable amount of time. This large a study is necessary because it has been noted that human reliability data may be statistically unreliable when obtained from the observation of too few trials to allow for estimates of performance variability (Cronback, 1947).

These field considerations fall into the category of evaluations of system reliability. The purpose of these studies have been three-fold:

1. To look for trouble areas (high probability of error) in the system.

2. To estimate the overall performance effectiveness.

3. To develop measures for personnel sub-system components.

In addition, the data obtained for personnel sub-system components has contributed to building human factors background in the data bank mentioned. This approach, with proper quantification, was necessary to convince system planners and designers that errors can be reduced by the application of human engineering techniques to the design of equipment (Swain, 1964).

A distinction can be made from this type of "evaluation" in the field from what may be termed "experimentation" in the field. The difference involves the opportunity to manipulate experimental variables. Manipulation of experimental variables can generally be considered to be inversely related to the degree of simulation of the system field configuration. Rabideau (1964) states that no manipulation is possible without disturbing the system's operation to some degree. *

2. Team Testing

One of the problems in assessing team performance has been the lack of adequate measuring techniques. Smode (1960) attributes this, in part, to not being able to specify task characteristics and general categories of behavior for working groups. The data gathered on social interaction in the group or team setting (Zajonc, 1965; Bavelas, 1950) have given some

* The field effort of this study is primarily an experiment rather than a field evaluation or attempt to determine system error rate or reliability. It is certainly possible that the data obtained could be treated in a systems reliability manner, or contribute to a general data bank, if indeed sufficient data is obtained to do this. The main goal however, is to determine the degradation of team variables with the imposition of the experimental variable of protective masks.

insight into what variables can be observed to assess team performance in the field. These studies, coupled with a task analysis approach (Kurke, 1961) involved team intercommunication behaviors.

Although the literature does not demonstrate a clear precedent for the observation of small team processes on maneuvers in the field, human factors reliability studies and commentaries do lend some guidelines for methods of quantifying measures, and data collection where several persons are performing simultaneously. Rabideau (1964) points out that the number and complexity of the dependent variables observed should be guided by the following points:

1. The observation should not affect human performance or inter-activities within the system. To the extent they do, the study is a less valid representation.

2. The number and nature of dependent variables will affect the treatment requirements (experimental design) and statistical reduction employed.

3. The reliability of field data recording process and devices affects the reliability of the obtained measures and the number of replications required.

4. Measurement requirements can impose personnel safety hazards, particularly when the field experiment involves a group's exposure to potentially hazardous forces. In this case certain measures may be deleted, made remote, or a safety monitoring provision required.

3. Quantification of Observations

Quantified measures taken in this area of human performance have traditionally been broken into:

1. Time readings

2. Error counts

3. Frequency data

4. Work load data

5. Motion dynamics

Time measurements can cover total time, sub-task times, or time

delays, but it is necessary to define all of these measures in the context in which they are being employed. Error measurements have been encompassed by one of four general categories:

1. Performance of a required action incorrectly.

2. Failure to perform a required action.

3. Performance of a required action out of sequence.

4. Performance of an action not required.

Rabideau (1964), in discussing error measures, adds the qualifying comment that the error must have a significant effect upon the system. The definition of such "significance" of impact could lead to "splitting hairs" in evaluation efforts.

A measurement developed for field use has been a frequency count of conditions which represent personnel safety hazards. The measure seems to be peculiar to field studies, and has been outlined into four categories, or types of error:

1. Safe personnel error: One which will not result in mission degradation or personnel injury.

2. Marginal personnel error: This will degrade the total to some extent, but can be counteracted.

3. Critical personnel error: Will degrade the mission unacceptably, causing personnel injury.

4. Catastrophic personnel error: Causes severe degradation of the mission, with possible loss of life to personnel.

4. Field Data Collection Techniques

The primary data transducers for field studies have traditionally been human beings. If communications are required between test subjects, the trained human observer attempts spot evaluations and subjective judgments, a type of data collecting that hardware cannot do. Because of versatility, human observers have been relied upon heavily in the field. It is usually anticipated that observers, especially contractor observers, will require considerable background training in observing and recording data before their skills would be satisfactory for a large scale test program. The main benefits of an observer can be summarized as: ease of movement, capability of recording complex behavior, and general versatility. Some potentially serious observer drawbacks are inattentiveness, speed, and

accuracy in recording. Also, an observer's training time, operating and field maintenance cost must be considered.

Tradeoffs with respect to a selection of data transducers should be considered, however. Recent innovations in technology have introduced refinements in recording equipment that have aided field measurement. The use of videotape, color film, and audio recording has been used intermittently to date, but some advantages have been recognized (Smode, 1960):

1. Film and videotape, upon analysis, can emphasize dynamic aspects of a process not readily seen on site.

2. They are permanent and have versatility with respect to the time continuum.

3. They can be used when additional verification of interpretations of data is desired.

4. They are particularly useful where safety hazards are present.

If the field of observation is not localized, a video or audio recorder may not be applicable, but can be most suitable where the trials are at a fixed duty station.

5. Protective Equipment

Research on design and field testing of protective equipment has highlighted certain aspects of the protective mask. A consideration of a few of these can provide a background against which variables relative to team interaction may be evaluated in the performance of a team in an operational environment when using a respiratory (oronasal-ocular) protective device.

The decrement in vision has been largely in the peripheral area (Shephard, 1962; Goldberg, 1965). In most protective masks there is a significant reduction in the physical visual field. Acuity and depth perception have been noted to be affected very little within the visual field. Shephard (1962) does report a reduction in binocular vision of near-low objects; however, and Goldberg (1965) reports a further reduction in the visual field due to the possibility of eyelens fogging at certain temperatures and humidities. Visual difficulties during attempts to use standard military optical devices while wearing a protective mask have been reported (Wulfeck, 1958). The problem here was due to eye relief distance. Military optical devices are designed for use by the unencumbered eye, and the eyelens separation degrades the performance of a man trying to use optical equipment with a mask. The adequacy of visual contact with outside surroundings can further be impaired by lens distortion in the peripheral area, glare,

and possible eye irritation. The reduction of the peripheral field has been studied statically, leaving the dynamic aspect relatively unidentified. Shephard (1962) and Goldberg (1965) mention that compensation will obviously be beneficial in a dynamic situation. Goldberg (1965) terms this the "mobile visual field"--the visual field obtained with compensatory head and body movement. Although peripheral vision is reduced it does not mean that an individual will not have 360° vision by adroitly using head and body movements. But, in order to maintain an adequate "mobile visual field" a man may need excessive head movements which may affect his performance. The effects could relate to the equipment, or in a group team task to his perception of communication signals. As noted by Gruber (1964) the use of hand signals is very important in combat effectiveness. A protective mask may hinder a team member's reception of this type of communication.

The inspiratory resistance offered by the mask has been reported to attenuate the transmission of sound energy due to this type of barrier effect. Alteration of pitch and quality, as well as the addition of respirator noise to the voice signal have been noted by Goldberg (1965). Speech distortion and attenuation may seriously degrade coordination that depends on verbal (oral) commands. The M-17 mask has been reported to transmit voice signals in a somewhat squeaky, higher pitch than normal speech. (Dunlap and Associates, 1963).

Further studies of a field nature have offered additional information regarding adaptation to protective equipment. In an exercise conducted at Ft. Hood, Texas ("Retort," 1964), it was reported that much difficulty was experienced in radio transmission while wearing the M-17 mask. Nearly every transmission had to be repeated. The report suggests a built-in microphone attachment be devised, if the M-17 mask is not adequate for personnel who are required to use a radio for command and control.

Tank and personnel carrier drivers discovered they could not see due to the limited peripheral vision. They reported the eye pieces in the M-17 as being too small for a proper job of driving. This did not apply to all tasks, however. A general conclusion from the report was that a well trained unit can perform its combat mission for a limited period of time (3 days) under chemical conditions with only minor loss of efficiency.

Training with protective equipment is mentioned in several reports as rather crucial to the success of a unit mission under CBR conditions. Army Field Manual 21-48 (1964) suggests an integrated CBR approach by having the men masked for half a day during normal on-the-job field training. Part of the acceptance of the mask by the men involves confidence in its protective ability and both physical and psychological adaptation to it, which only experience can give. The field study at Ft. Hood ("Retort," 1964) concluded that a properly trained small unit leader can decrease casualty

assessment by exercising proper CBR discipline. Shephard (1962) further notes that, on self-paced tasks, endurance is poorer with the protective mask, but with proper training, adaptation can be facilitated.

Observations collected by Dunlap and Associates personnel (1963) during infantry field exercises indicated marked differences in speed of movement when troops were masked, especially when moving through brush. Shephard (1962) reports that psychomotor performance is generally unaffected by a protective mask, and in simple manual tasks the degree of degradation is statistically insignificant in most cases. In a study conducted by Dunlap and Associates, Inc., (Gruber, 1964) a comparison of performance between masked and unmasked troops indicated significant differences in firing response time of individual soldiers from a kneeling position after securing cover. Their rate of fire was also significantly less, but accuracy showed no apparent differential effect. Dunlap and Associates' 1965 observations at Camp Pickett, Va. on individual behavior in the face of simulated machine-gun firing indicated a greater tendency of masked personnel, crawling "under fire," to break and run, at least in training exercise, than unmasked soldiers.

A number of physiological responses have been studied and noted under conditions of wearing masks when performing various tasks. In Exercise "Retort" (1964) it was reported that wearing the M-17 mask for four hours produced drowsiness and headaches in about 10% of the men. It was expected this was due to difficulty in breathing and headstraps, especially to the tightness of helmet straps. General discomfort has brought additional stresses of various types on the men, decreasing the individual soldier's acceptability of the mask. But as previously discussed, training can be crucial in this regard. Shephard (1962) discusses a few reasons for possible distractions due to overall discomfort. A few of these are pressure on the face, thermal effects, water vapor condensation, and possibly claustrophobia. The tolerable and optimal levels of temperatures and humidity for various levels of motor activity have not been determined, but discomfort due to these variables may be expected to adversely affect proficiency. There is, of course, the possibility that at extremely low temperatures the warmth of the mask could be a benefit.

A possible problem area for men performing manual tasks in the field, relates to the breathing resistance of a mask. If breathing is hampered, this can add to the physiological burden, and shorten the time to exhaustion. Goldberg (1965) notes that change in the basic rhythm of breathing alone is subjectively objectionable, even when no particularly difficult physical activity is being performed. Goldberg (1965) quotes several studies that have investigated various breathing impedances and various physical work loads. It is shown that higher resistance face pieces result in insignificant differences in pulse rate at the same level of work.

Pulse rate increases in a general linear way with linear increases in work load, with or without the masks. With the increased resistance mask, respiratory rate increased, but minute volume decreased about 10%. Even soldiers who worked to exhaustion with the high resistance masks did not show that their physiological limits were significantly affected. The subjects were, however, more aware of the higher resistances.

The basis problem in mask design has been to develop an effective chemical agent protective system that will cause only minimum decrement in task performance. Task performance alone, however, is not sufficient for mask evaluation. Its subjective acceptance needs also to be assessed to assure that the mask will be retained through combat conditions until needed. Further, there will always be a need for training to overcome the initial objection to wearing any facial or head covering device.

Summary

The above background literature served as a broad guide in proceeding to the next phase of development: the specific design of an initial structure of team performance measurement and of a field test of this intended practical structure. This phase of design effort is described on the following pages.

B. Initial Observation Test Design

1. Following is preliminary design for the observation aspects of a field evaluation test of team performance, as initially developed during the first quarter of this study. It illustrates the extent to which a standard test technique is capable of delineation on paper prior to field observation experience.

a. Representative Tactical Task

The team task for initial field observation should exhibit these features:

- Localized, compart site of team activity.
- Small number of team personnel.
- Tactically significant task.
- Brief, repeatable activity.
- Validity of task performance with gas protection.
- Standard tactical criteria for accuracy and procedures.
- Diversity of modes of team interactivity (visual, physical, oral).

The sample task used in initial test design has been the emplacement and preparation for firing of a 155 mm Howitzer. However, this part-task is brief and involves too many men (10) for ease of observational design. The "Little John" rocket artillery task was tentatively suggested for initial field observations as meeting the test design criteria. The diverse activities of its small team are localized at a compact emplacement site. It is a repeatable, high-priority tactical mission with emphasis on both speed (10 minutes) and accuracy of team coordination. The alternative might be the 105 mm Howitzer which is more readily available, and which also requires a relatively small team. (This was the system finally chosen, largely because of availability).

b. Performance Measures

Protective equipment testing shows an uneven history of significant decrements in gross time and accuracy. Individual variations in physiological and psychological behavior while wearing masks are wide-ranging. Team testing adds to these variables the complexities of inter-member compensations and coordinations.

This test design, accordingly, should place emphasis not on gross differences but on a finer analysis, in some depth, of the comparative patterns of team activity and interactivity with and without masks. Sub-tasks of the test activity can be separately analyzed for changes in sequence, in density of team participation, in modes of inter-member coordination, and in number of critical incidents with and without respiratory protection. Time and accuracy measurement can be segmented into sub-task times and observer evaluations of separate team activities.

Changes in the pattern of intercommunication among a team's members must be monitored in particular since protective masks and hoods do directly alter vision, hearing, speaking and other signaling. An altered pattern of team efficiency due to restricted breathing and movement might be a secondary object of monitoring and measurement.

c. Hypotheses for Initial Testing

These were the tentative hypotheses about anticipated differences in the performance patterns of teams with and without chemical protective gear prior to field activities (based on initial National Guard observations):

Total Time (Quantity of Activity)

Team with protective head gear will perform slower than team without the gear; i.e., will take more time to accomplish the total task.

(1) $T_w > T_{w/o}$

Overall Accuracy (Positive Quality of Activity)

Team with protective head gear will perform with less precision than team without the gear; i. e., will compile fewer accuracy points on the objective test ratings.

(2) $A_w < A_{w/o}$

Incidence of Deviations

The incidence of unprogrammed activities (delays, accidents, deviations from intended procedures, etc.) will be greater for the team wearing protective head gear than for the team without it.

(3) $I_w > I_{w/o}$

NOTE: Any "unscheduled critical incident," i. e., an obvious repetition of oral or visual signal, a reversal of efficient sequence of procedure, fumbling, delay or repetition of a sub-task will count as a deviation.

Density of Team Activity (Efficiency of Activity)

Members of team that wear protective head gear will individually be active for a greater proportion of their total task time than members of the team with the gear; i. e., masked team members will work with fewer rests.

(4) $D_w > D_{w/o}$

Visual Interactivity (Efficient Visual Coordination)

Team with protective head gear, in coordinating its member activities, will use more broad visual signaling than team without gear; i. e., a greater number of observer-detectable visual signals.

(5) $V_w > V_{w/o}$

Oral/Aural Interactivity (Efficiency of Oral/Aural Coordination)

Team with protective head gear, in coordinating its member activities, will use less oral signaling than team without head gear; i. e., a fewer number of observer/audible oral/aural interchanges.

(6) $O_w < O_{w/o}$

Corollary Hypothesis

(6a) $O_{w-h} > O_{w/o-h}$

Team wearing protective hood with mask will complete fewer effective oral signals (i. e., with more repetitive oral deviations) than team wearing only masks.

NOTE: Hypotheses (5) and (6) are not predicting that there will actually be more visual signaling and less oral signaling activity by the masked group. They are concerned with the observable and effective activity. In other words, (5) is predicting that visual coordination will contain more gross signaling motions, detectable by the visual interactivity observer, in place of more subtle visual cues among team members possible when they are not masked. Both (5) and (6) are concerned with number of effective acts of coordination, i. e., even though masking might produce more frequent repetitions of oral attempts of coordination, all unheard or unintelligible attempts would count as incidents of deviation, and only the oral/aural exchange resulting in an interactivity would be counted to test hypothesis number (6).

d. Detailed Measurement Plan

During experimental field testing, the following measurements should be attempted to test adequately the paper and pencil hypotheses put forth:

- Actual sequence of team sub-tasks
- Number of visual, oral and manual team coordination cues
- Quality of performance of each team sub-task
- Time to perform each major team sub-task
- Density of team's cumulative activity during each sub-task
- Number of critical incidents and accidents
- Subjective rating of performance by teams
- Tactical accuracy of team performance
- Total time of team performance
- Comparative indications of physiological energy expenditure (via telemetered heart-rate)

From these measurements of each team's test runs, both with and without protective masks, an analysis can be evolved in depth for any

consistent or significantly different patterns of team performance under the two conditions. These pattern analyses -- together with cumulative time, accuracy, incidents, and other totaled and averaged data--should serve to prove, disprove or modify the tentative test hypotheses.

e. Data Collection Plan

Upon selection of tactical task to be observed and the field site for observation, an Observational Unit must be assembled and instructed in its respective data collection duties. This field testing team would require three basic personnel types in addition to the participating tactical teams assigned. Following are the duties, as detailed in the initial plan, of these personnel:

- Monitors
- Recorders
- Observers

The Monitors will include:

- Dunlap personnel responsible for the standardized test
- Army project coordination personnel
- Supervisors of tactical and technical teams

The Recorders will include:

- Visual recording team (two film or two TV cameramen)
- Oral recording team (audio taping and pick-up technicians)

The Observers will include:

- Visual Interactivity Observer (VIO)
- Oral Interactivity Observer (OIO)
- Team Events Observer (TEO)
- Test Umpire

Visual Interactivity Observer (VIO)

Observation and collection of all visual data will be this man's responsibility. He will:

- Position cameras for best visual recording of test activities

- Observe each team sub-task for use of the visual mode of team coordination

- Record on the visual interactivity format each use of the visual mode of coordination

- Identify and count each critical incident or deviation in a team's visual interactivity

- Comment, via audio recorder or written notation on any visual team data demanding interpretation

Oral Interactivity Observer (OIO)

Monitoring and collection of all oral/aural data will be this man's responsibility. He will:

- Position microphones for best aural recording of test activities

- Observe each team sub-task for use of oral mode of team coordination

- Record on the oral interactivity format each use of oral mode of coordination

- Identify originator, intended hearer, and probable intelligibility of each oral intercommunication

- Identify and count each critical incident or deviation in a team's oral interactivity

- Comment, via audio recorder or written notation on any oral team data demanding interpretation

Team Events Observer (TEO)

Evaluation of the quality of team performance will be the on-site responsibility of this man. He will:

- Observe the performance of each major sub-task

- Rate each performance subjectively on his format

- Identify and comment on any sub-task elements rated "poor"

- Count and describe any observed critical incidents, particularly in sequence and efficiency of team sub-tasks

Test Umpire

The constant monitoring of actual test runs for any violation of the constraints of test design will be the responsibility of this man. He will:

- Identify the commencement of each team sub-task from the start activity described on his team task diagram

- Inform the members of the Observational Unit of the conclusion of one sub-task and commencement of the next

- Match which and how many team members are actually performing each sub-task against the SOP groupings on the team task diagram

- Make notation or taped comment on any deviation, such as skipped or reversed sub-tasks, laggard or over-active personnel, and new or different team procedures

NOTES: (1) In initial field observations, if conducted on a non-interference basis with tactical training, the Test Umpire will merely note all deviations from desired test control (i.e., an over-eager team member helping on sub-tasks other than those assigned to him). In later standard runs of the team test, the Umpire could declare a run with major deviations cancelled, to avoid statistical distortion of cumulative data.

(2) For facilitation of observation, each tactical team member will wear his number prominently displayed on chest and back.

f. Data Analysis Plan

Audio/Visual Recording

This test design assumed visual and oral recording of field runs by camera and tape recorders as an integral part of the Observational Unit. This allows a small team of observers to monitor all members of the tactical team without attention to tedious manual timing and detailed counting of sub-task activities. All such specific quantified data can be derived from later repeated monitoring of the audio/visual record of runs.

Multi-tracked tape recording would be preferable, if available, for synchronized tactical team audio and observer commentary. Television tape recording would be preferable to film because it allows:

- Extensive field recording, with subsequent erasure of poor or discontinued runs and reuse of the tape
- Immediately available visual records for rapid training of the observation team in needed interpretative commentary and notation
- An overlap of initial data analysis with field runs still in progress in case modifications or reruns of an assigned team are indicated for useful analysis records

2. Design of Field Task for Observation

The tactical activity design for the observation phase as approved by CONARC was of an "RSOP" (Reconnaissance, Selection and Occupation of Position) activity: 105 mm howitzer sections conducting modified "deliberate occupations." A "deliberate occupation" can be defined as a planned occupation of an area by all elements of a howitzer battery. The occupation is preceded by a reconnaissance party in order to determine howitzer firing positions, installation sites for all the battery elements, routes of march, defensive plans for the battery, etc.

Just as the observer tasks were specified during the design phase, so also the basic field procedures of the members of the 105 mm howitzer sections who participated in the occupation were partly structured. Each section was to complete four major actions: uncoupling the howitzer, preparing for actions, firing and march order. To facilitate the data collection methods, the major actions were divided into five sub-tasks. A brief summary of each sub-task is given in succeeding paragraphs. The flow of the team activities are described in the team task diagrams.

The howitzer team was to mount into the 3/4-ton truck (prime mover) in the assembly area under the control of the Chief-of-Section. Each member of the team would be equipped with the following clothing and equipment: fatigue uniform, field jacket, helmet w/liner, M56 harness with ammunition pockets, canteen, intrenching tool w/carrier, and rolled poncho, the M16 rifle slung on the right shoulder and the M17 protective field mask carrier worn on the left leg. The team members would don the M17 mask in the assembly area prior to a masked observational run. All members of the team would wear numbered vests to identify their duty positions more easily during observation. Once team and data recording devices were ready, the Team Umpire would notify the Chief-of-Section to have the truck

driver approach the Starting Posts.

Sub-Task Summaries

The task flow "Data Sheets" contain a series of discrete actions for observation and evaluation under five major divisions described in general terms below:

Sub-Task 1 - Uncoupling

The Chief-of-Section dismounts from the prime mover to guide the vehicle through the starting posts and to the marked firing point (see Figure 2). At the firing point, the Chief-of-Section gives the command to halt and dismount. The section dismounts and uncouples the howitzer. The gunner and assistant gunner immediately prepare the fire control equipment on the howitzer in order to align the gun in a direction of fire. The remainder of the section spread the trails of the howitzer and begin unloading the truck. An aiming circle, placed to the left front of the howitzer, is used by the gunner as an initial aiming point. Sub-task 1 is completed when the gunner announces "Aiming Point Identified, Sir." A typical time to complete Sub-task 1 is approximately one (1) minute.

Sub-Task 2 - Prepare for Action

Sub-task 2 follows Sub-task 1 without a time gap. The aiming circle instrument operator announces a deflection to the gunner. The deflection is set on the panoramic telescope by the gunner. The gunner directs the cannoneers in the shifting of trails by using oral and visual signals. The howitzer orientation is checked by the aiming circle instrument operator after the gunner announces "Number ___, Ready for Recheck, Sir." This process is repeated until the howitzer is pointing in the direction of fire. Simultaneously, while the howitzer is being oriented, several cannoneers complete unloading the truck and prepare the ammunition for firing. After the howitzer is oriented in a direction of fire, the aiming posts are placed out at a common deflection, approximately 50 and 100 meters from the howitzer. Spade holes are dug to facilitate firm seating of the trails. The site to the piece mask range is determined by the Chief-of-Section. Announcement of the site to mask indicates the completion of Sub-task 2. A typical time to complete Sub-task 2 is approximately four (4) minutes.

After the completion of Sub-task 2, a safety check is conducted by the safety officer to insure that the howitzer is oriented in the proper direction of fire. If any discrepancy occurs, the howitzer is reoriented until it is considered safe to fire. The time required for this safety check is not to be recorded or considered for data analysis. Upon completion of the safety check, a fire mission is given to the howitzer section. The fire

mission originates from a battery officer at the battery position. It is processed by the Fire Direction Center, who sends the computed data to the battery executive officer.

Sub-Task 3 - Initial Firing

A sequence of firing commands are given to the howitzer section in order to prepare the howitzer for firing. The firing commands contain the following information:

- Warning Order
- Type of Shell
- Charge to be Fired
- Fuze Action
- Which Howitzers will Fire
- Deflection
- Elevation
- Order to Fire

During this sub-task, one round of ammunition is prepared to fire. The howitzer is oriented in both deflection and elevation and is loaded. Prior to firing, the safety officer checks to insure that all the elements of the firing commands have been correctly carried out. When the safety officer states "Safe to Fire" the command to fire is given by the executive officer. Sub-task 3 is completed when the Number 1 Cannoneer announces, "Bore Clear, Number ___, Sir." A typical complete Sub-task 3 is approximately one-and-a-half (1.50) minutes.

After the round impacts in the target area, the battery officer states the necessary shift in deflection and range in order to effectively destroy the target. This information is processed by the FDC and the firing commands are relayed to the battery executive officer.

Sub-Task 4 - Fire for Effect

A second sequence of firing commands is given to the howitzer section. Two rounds are prepared for firing. The howitzer is reoriented by a new deflection and elevation, then is loaded. The safety officer again checks the howitzer and states "Safe to Fire." The battery executive officer

gives the command to fire the first round. The second round is loaded immediately. The deflection and elevation is rechecked by Gunner and Assistant Gunner and the command to fire is given by the Chief-of-Section. Sub-task 4 is completed when the Number 1 Cannoneer announces, "Bore Clear, Number ___, Sir." A typical time to complete Sub-task 4 is approximately one (1) minute.

Sub-Task 5 - March Order

This sub-task is initiated upon the battery executive officer's command of "March Order." During this sub-task all the howitzer equipment is placed back on the truck in the fastest way possible by the entire howitzer crew. The howitzer trials are closed and the gun is coupled to the truck. The section mounts up on the command of the Chief-of-Section. This sub-task is completed when the truck departs from the firing site. A typical time to complete Sub-task 5 is approximately two-and-a-half (2.50) minutes.

Design Summary

General design was approached through investigation of the background problems and through survey of the pertinent published literature. Specific design was then developed in a consistent, logical manner by detailing the steps for feasible field testing of team evaluation techniques. Selection of a howitzer task allowed descriptive division of the intended observation into sub-tasks. With the conclusion of this initial design and the simultaneous provision for its field testing; the study moved into preparations for that phase, described in Section III.

SECTION III

FIELD OBSERVATIONS

A. Background

The design of a plan for field observations, as conceived during first quarter study, was reviewed and approved at Edgewood Research Labs in October, and the availability of various possible test sites and team tasks, to meet the design's criteria, was determined in November. A statement of "Field Observation Requirements" was then drafted for use by the Contract Project Officer in submitting a formal request, from Edgewood to CONARC, for the field observation site, personnel, scheduled exercise, and necessary support.

The feasibility of such a request was informally discussed at CONARC in November, formally submitted during December, and was authorized, in January, for observations to take place at Ft. Bragg, North Carolina. A visit to Ft. Bragg on 28 January 1966 established liaison, located the requested field exercises on the training agenda of the 82nd Airborne Division, and scheduled the observations for the week of 21 February 1966.

B. Preparations

With excellent cooperation from the 2nd Battalion, 321 Artillery, 82nd Airborne Division, under Lt. Col. Joseph Baker, a standard RSOP exercise in deliberate occupation was planned for the designated 105 mm howitzer battery, with the additional objective of achieving a balance between the structuring needed for an effective performance measurement environment and the non-interference requisite to maintaining customary team behavior in the field environment.

To maintain necessary realism, the following factors were included:

a. Use of moderately unpracticed teams in actual training.

b. Emplacement at an actual field site on the firing range.

c. Full tactical uniform, equipment and live ammo.

d. Complete emplacement, live firing, and march order sequence.

e. Field training SOP (Standard Operating Procedure).

To establish a test structure for objective performance measurement,

these factors were included:

a. Equal manning (eight-man team) for all six sections of the battery.

b. Same man at same position for pairs of trials, under both conditions (with, vs. without masks).

c. Observation of each team under both conditions by the same set of observers and recorders.

d. Reversed sequence of trial conditions for half of test teams, in alternation.

e. Standard orientation briefing for all teams participating.

f. Three rounds, one initial and two for effect, to be fired at each trial run of SOP task.

Several minor problems in structuring the test while maintaining field realism were encountered and resolved as practically as possible. For example, numbered vests and telemetry gear worn by the men were attached so as to cause minimum interference. Also, cameras, observers, and recording equipment were situated well outside the radius of normal team activity; and observers were dressed inconspicuously and cameras fitted with long-range lenses to minimize disruption. When a test subject had to be used again to fill a later team, he was always assigned the same position and the duplications recorded.

Preparations were coordinated among the Project Director at Dunlap and Associates, the Contract Project Officer at Edgewood and G-3 Headquarters of the Organization and Training Division, 18th Airborne Corps at Ft. Bragg. The responsibilities of each of these participating organizations were assigned in detail some weeks ahead, and the exercise scheduled for the site beginning at 0800, 21 February 1966.

C. Field Exercises

1. General On-Site Activity

Monday, 21 February - Observation equipment was brought to the site and tested. TV platforms were erected, electrical power brought to site, the tent for telemetry equipment provided, and locations for start of observation and for emplacement staked out.

Tuesday, 22 February - All observation techniques were reviewed, observers and camera crews briefed, and equipment checked out for proper operation. Telemetry, TV and audio recorders functioned perfectly.

Wednesday, 23 February - All military participants, observers, and equipment were ready for observation on site by 0800. Battery "A" personnel were given the standard orientation to the procedure and purpose of the field exercise. Two dry runs (without firing) and nine test runs were completed during the day, with only one run being disqualified by the Test Umpire (team believed it a dry run).

Thursday, 24 February - Rain throughout the previous night and the entire day forced postponement of exercises until Friday.

Friday, 25 February - All observers, participants and equipment were on site and ready for activity by 0800. After a 45-minute delay for verification that Medic requirements were met, exercises commenced with test run #10 on the schedule of official observations. The hooded condition (M-17 mask fitted with its standard hood) was observed for the two final sections of Battery "A." The first two sections of Battery "C" were briefed and observed in afternoon runs 16 through 19, but their performance of the task diverged so markedly from the SOP of both Battery "A" and the prepared OSD of the task that a tentative decision (later confirmed) was made to disqualify these final four runs from the set of data for statistical analysis. Remaining ammunition was expended in brief practice round exercises, and the exercise was completed at 1600.

2. Detailed On-Site Activity

Activity on the Ft. Bragg site can be reported, for this study's purpose of gaining insight into feasible field team evaluation, by each discrete technique of observation attempted during the field experience. Each of these techniques will be described as to implementation, procedure, and general difficulties encountered:

- Trained Observers/Test Data Sheets
- Television Recording
- 16 mm Film Recording
- Recorded Verbal Commentary
- Telemetry of Heart Rate
- Supplementary Film Recording
- After-task Interviews (Individual Subjective Reports)

(A fuller analysis of the resultant data collected is presented in Section IV, and the advantages, disadvantages, and recommended adaptations of these

techniques for field team observation are discussed in detail in Section V).

The overall technique used during these field observations was to place coordination of all participating military personnel, both participating teams and supporting FDC, under an Executive Officer, and all observers, monitors, technicians, and recording devices under a Test Umpire. Together, they coordinated the field activity. The Executive Officer coordinated activity at the FDC with that at the emplacement. The Test Umpire controlled the sequence of trials. His verbal and visual commands synchronized the starting, timed progress, and ending of all observation and recording of each run by number. An official trial number was recorded manually, visually, or aurally by the various observers and equipment at the beginning of each observation run. The starting and running times were continuously inserted during the run into the upper left corner of the videorecording image from the Umpire's electric running time meter by an off-site input from a TV camera.

a. Trained Observer #1: Test Umpire

Implementation - The Test Umpire was equipped with the following observation/collection aids:

- Test Umpire Data Sheet
- Portable Audio Tape Recorder
- Chalk Board, Chalk and Eraser
- Running Time Meter (seconds and minutes)

Procedure - The Test Umpire first identified the particular trial run and mask condition coming next on his chalk board for recording by TV and film cameras. Next, he checked the readiness of all personnel and equipment. Then with a standby command to alert all personnel and to start the videorecorders, he gave a manual signal to the truck driver to bring the test team into the observation site. His simultaneous signal "Start Run #   ," recorded on the TV audio track, served to synchronize the starting of his running time meter with the starting of the two heart rate recorders by the telemetry technician.

Next, the Umpire followed his Data Sheet's task flow to detect any radical reversal of subtasks or failure of team members to perform in the generally prescribed manner. Also, during the run, he noted the general weather conditions, the test environment, and his comments on a "Title Page" for each observed run: The Test Run Identification Sheet. On his audio recorder, he made specific comments, for later transcription, on any atypical behavior which might distort the objectivity of the data from that

trial run.

Finally, at the closing of the tail gate of the truck and first departing motion of the truck, the Umpire declared the run ended. He stopped the running time meter, reset it to zero, and erased the run number on the chalk board, preparatory to updating it for the next run.

Difficulties - The primary task of the Umpire, test coordination, was successful. In only one instance was there confusion over the sequencing of runs--when howitzer section 4 became team #3 in the observation sequence, and section 3 followed as observed team #4, temporarily confusing identification. Test Run #2 had to be disqualified because the participating team had received the erroneous impression that this was another "dry" (or practice) run. Synchronization of starts and stops was successful.

The secondary task of monitoring the team activity for departures from assigned duties and sequences on his Data Sheet (a simple Operational Sequence Diagram) was made doubly difficult by the lack of experience of the available teams and by the cold, raw weather. The observed teams relied upon their individual chief-of-section for task procedure, rather than on training in standard procedures. Thus a given team's procedure basically matched the Data Sheet, but was usually too dependent on the chief's personal directions to follow systematically by the SOP. The temperature was below freezing during some runs on the morning of 23 February, and the wind quite gusty, making note-taking by stiff or gloved hand tedious. The portable audio recorder, kept warm inside the Umpire's coat pocket, functioned satisfactorily and allowed for more extensive commentary for later transcription that would have been impossible to write on the scene.

b. Trained Obsever #2: Team Events Observer

Implementation - The Team Events Observer was equipped with the following observation/collection aids:

- Team Events Data Sheet
- Clipboard
- Portable Audio Tape Recorder
- Microphone with trailing cord to provide commentary directly onto TV tape
- Audio-Video playback equipment

Procedure - The Team Events Observer positioned himself to the left rear side of the emplacement site, close enough to provide ample view of the tactical team. As a test run proceeded, he moved within a range of several yards around the area observing the various individual activities taking place. The Data Sheet provided him with a task flow which aided his detection of any deviations from a normal operational sequence by the personnel performing the task. As the task progressed, he would check off each sequence and briefly note any deviations. At the end of a sub-task, he would subjectively rate the overall performance of the tactical team, relying on his experience as an Army Reserve Officer for judgments of team coordination, speed, safety, and smooth accomplishment. This observer used a portable audio recorder for comments on deviations and critical incidents. The recorder was also used for more general commentary for later transcription during data reduction. At the end of a data run, the Team Events Observer had access to the TV van where he immediately viewed a playback of the entire run on a TV screen. As a self-check he again rated the quality of each sub-task, as he recorded additional commentary directly onto the tape's extra audio track.

Difficulties - The Data Sheet used by the Team Events Observer was designed for a standard 105 mm howitzer section; unfortunately, the tactical teams observed had not functioned together long enough to adopt such a consistent basic operating pattern. Each team required close direction from its Chief-of-Section so that sequence and detail of activity varied, unavoidably, from the outlined task flow. Although overall quality could still be rated, it was difficult to follow the differing sequence of operation from team to team. As a consequence of the low temperatures, it was also difficult to record observed deviations by hand on the Data Sheet. The portable audio recorder was thus used constantly during the first five data runs, but malfunctioned during run six. The Team Events Observer, therefore, replayed the videotape of the next several runs after the conclusion of each and recorded his running commentary directly onto the videotape. This interim replay became time consuming, and it was found that the commentary could be satisfactorily placed on the tape during a run by use of a hand-held mike with a long cord trailing the observer onto the field from the truck. Aside from these minor difficulties in recording, the monitoring technique employed by the Team Events Observer proved successful.

c. Trained Observer #3: Oral and Visual Interactivity Observer

Implementation - The Oral-Visual Observer was equipped with the following observation/collection aids:

- OIO and VIO Data Sheet

- Clipboard

- Portable Audio Tape Recorder

- After-Task Questionnaire Formats

Procedure - This observer's primary task was to detect, count, and classify all communications attempts, visual or verbal signaling, made between members of a team during an observed run at the field site. By referring to his Data Sheet, the observer could anticipate certain signals, those required for SOP performance and coordination of the task, and could classify them as "scheduled communications" when they occurred. Communications, verbal and visual, that were specifically described in the 105 mm Howitzer Manual fall into this category.

All other interactivity was counted and classified in either of two additional categories: "unscheduled" and "repeated" signaling. If an additional, or "unscheduled," communication was observed, it was noted in the appropriate column of the Data Sheet as "manual" (visible) or "verbal" (audible), and the intent, content, speaker and hearer, or other descriptive data were recorded into the observer's portable audio tape recorder. If either a scheduled or unscheduled communication was repeated by a team member, it was tabulated on the Data Sheet as a "repeat," and described on the recorder as being caused by an ineffective initial signal, a performance deviation, or whatever. All of these counts and classifications were kept separately for each sub-task of the run as it progressed; there being definite sub-divisions of the task on the Data Sheet itself.

At the completion of the final sub-task, "March Order," the team dismounted from the departing truck at a debriefing site approximately 100 yards away where the Oral-Visual Observer then proceeded to his alternate field assignment--the administering of the After-Task Questionnaire. The format he passed out to the men was either With Mask or Without Mask according to the condition under which they had just performed the test task. This same procedure was used following all runs, except the two in the hood condition for which no debriefing was held.

Difficulties - The problem of each team's "modus operandi" varying from the Data Sheet was particularly troublesome in this observer's attempts to monitor both oral and visual communications. Although the Data Sheet was relatively well laid out, and though behavior patterns of the various howitzer teams were not critically different from its sequence, the additional skipping through and page-turning caused much loss of time and attention from the main task of observation. Some activities thus slipped by unnoticed. This was later confirmed when viewings of the videotapes by the same observer resulted in appreciably greater intercommunications counts.

On-site, the division of time among audio recording, turning Data Sheet pages, placing check marks, and observing was particularly detrimental to detection of unscheduled intercommunications. By turning up the volume of his recorder, the observer was able to record most of the verbal unscheduled signals and to classify them later at his leisure from the audio playback. This allowed him more time for monitoring manual signaling during the runs. The more general problem of overhearing quietly spoken commands or remarks was handled by classifying them by context; the actions of the men in response usually aided in their identification as scheduled, unscheduled, or repeated "verbals." Assigning and giving a discrete count to the informal discourses among the small working groups was often arbitrary, of necessity. But seldom did these incidental remarks appear to represent critical communications for the coordination of the task.

d. Television Recording

Implementation - A mobile TV recording system was brought to the site in a small van by I. T. V., Inc. of New York City. The equipment included:

- Three Television Cameras and Amplifiers
- Two Videotape Recorders
- Two Audio Channels and Microphones
- Several Hours of Videotape
- An image insert generator, miscellaneous support equipment and lenses.

Procedure - Two television technicians from I. T. V., Inc. operated the equipment during field activity, one running the camera assigned to "close-up" views of team activity and the other monitoring the two recorders and the video chains inside the small van. Each morning the TV cameras were checked out prior to scheduled observation time, a test sequence recorded and played back, and the "wide angle" camera fixed on its pedestal to include the entire area of team activity at the emplacement site (See "Site Layout Diagram," Preface, Figure 2).

At the Umpire's ready signal for a trial run to get underway, the cameraman focused on the chalk board identification of the run's number and condition, and the recording technician started the two recorders. As the towed howitzer pulled onto the emplacement site and activity progressed, the cameraman followed the critical individual or small group actions on the close-up camera while the fixed wide-angle camera recorded the entire

field of team activity including howitzer and truck. At the Umpire's signal for end of run, the recorders were stopped, and the ending footage of the reel was read and noted on a log of the location of each run on the recorded tapes. A microphone placed beside the team had recorded all verbal exchange onto a sound track alongside the video.

Difficulties - The best view of the 105 mm howitzer team task had been determined to be from the right side, unblocked by either truck or ammo activity. Unfortunately, layout of the range required firing to be toward the southwest, causing the right-side camera to face into the morning sun from the north. A raised platform installation for assuring clear camera lines of sight also controlled lens flare, but the view was still backlighted from this angle, rather than frontlighted. Ideally, the wide-angle camera should have been to the southeast.

The numbered vests, worn by each man for his visual identification on the TV tape in later analysis, also offered some minor difficulty in staying tied down and in remaining legible on the wide-angle picture. Lack of consistency of procedure by the inexperienced teams observed made it impossible for the close-up camera to follow a scheduled sequence of critical activities. In a closer adherence to task flow, the Umpire could have "talked through" these activities in directing the cameramen by headphones to comparable close-up views of each sub-task for each trial run.

A gusty wind resulted in some recorded noise on the non-directional mike, even with it shielded. In general, all video recording was excellent, difficulties were minor, and the entire 19 runs of field activity were recorded, both wide-angle and close-up, on two four-hour and two one-hour reels of video tape.

e. 16 mm Film Recording

Implementation - Two Army film cameramen from Edgewood Arsenal brought the following equipment to record observations:

- Two 16 mm Aeroflex Cameras
- Full Complement of Accessories--Lens, Tripod, etc.
- A Supply of 400-foot Reels of Commercial Color Film.

Procedure - Film recording followed much the same pattern as videotape recording. The fixed wide-angle camera, however, was placed on a tripod on a raised pedestal to the south, or left-hand side of the howitzer, to take advantage of the front lighting for color filming. The second cameraman was assigned to take "roving" close-ups of a specified sequence of critical tasks, such as unpacking and fusing the shells while

wearing gloves or running out aiming stakes. The Umpire's chalk board was filmed for the beginning of each run, and then the wide-angle camera was run continuously during the activity at 24 frames per minute (sound speed, in anticipation of later addition of magnetically sound-scribed commentary). Film was changed between runs in a standard black cloth bag used for such field assignments.

During periods of inactivity of the teams being observed, the close-up cameraman took a series of film clips of the field observers, military monitors, site layout, and general orientations to the evaluative activity. These were for later incorporation into a sound film explaining the study's approach to field team observation.

Difficulties - The unexpected length of a few runs by inexperienced teams (up to 20 minutes) made it necessary for the close-up camera to "relieve" the wide-angle camera for short periods while a fresh reel of film was loaded, in order to maintain continuous film record of the long runs. Minor difficulties in getting an unblocked picture of some close-up tasks, such as the jacking of the howitzer from the front side, were largely resolved by an experienced and energetically moving cameraman. Complete coverage of the activity, even of minor interruptions, such as the safety officer's check, was necessary because of the danger of missing the recommencement of activity otherwise. Thus, total film footage during the observations amounted to an impressive, and expensive, sum of over 12,000 feet.

f. Recorded Verbal Commentary

Implementation - Provision for transcribing formal and informal running commentary by observers and monitors at the site was through:

- Four Audio Tape Recorders (See "Field Observation Equipment" for specifications).
- Two TV Audio Channels and Microphones
- Notebooks

Procedure - Each of the three observers carried a personal audio tape recorder on a shoulder strap and recorded comments at the moment of noting an event for which there was no formal classification on their Data Sheets. Their incidental qualitative remarks were also recorded as well as description of details or errors outside the clear view of the TV and film recording cameras. Commentary was flagged on the audio tape reels through verbal identification by run number and sub-task

for later transcription into typewritten items. Additional running descriptions of the runs were recorded by the Team Events Observer, at first by viewing a replay of the videotape after each run, and later by taking a microphone with him to his field position and commenting on each activity as it occurred. The TV tape thus carried one audio channel of the team's own verbal interactivity and a second, simultaneous input of observer commentary.

The remarks of other monitoring personnel on the site were also recorded on audio tape in some instances, and written notes were made in field notebooks to help later in coordinating in the proper sequence the data collected by different techniques.

Difficulties - The audio recorder used by the Team Events Observer gave some trouble, and informal commentary was insufficiently recorded from portions of early runs. The technique of formal commentary during a replay of the videotape after each run was found to be lacking in detail and took longer than was necessary for other between-runs preparations. With a change to the hand-held mike input to the videorecording system directly from the howitzer site, commentary detail and clarity was much improved, and the Team Events Observer was given a rest between runs. The most serious difficulty, only discovered afterward, was that both formal commentary and team verbal interactivity had been recorded on the same audio channel on the tape; the inputs from the hand-held mike and the omnidirectional mike near the team had been electronically mixed and recorded together, so that they could not be separately played back, even though a second, separate audio channel had been available on the tape. This made later analysis of verbal interactivity unnecessarily tedious and may have caused some verbal interchanges to go unnoticed. Wind noise effects were also bothersome on the omnidirectional mike.

g. Telemetry of Heart Rate

Implementation - The following equipment collected heart rate data from three members of each observed team:

- Three Individual Heart Rate Monitoring Systems
- Two Two-Channel Graphic Recorders
- Tent equipped with accessory gear and application materials for the outfitting of personnel with the sensor/transmitter elements.

Procedure - Three separate heart rate monitoring systems, described in full detail in a Supplementary Data Item, were developed prior to the actual observation phase.

A tent was erected each morning of the exercise by the participating howitzer battery to house the receiving elements of the system. This tent served a two fold purpose: 1) it provided protection from the weather for the receiving elements, i.e., receivers and recorders, and 2) it afforded protection to the subjects while they were being equipped with the system because of the cold temperature and gusty wind.

Prior to each observational run the Gunner, Assistant Gunner, and Number 3 Cannoneer reported to the telemetry observer. Each subject removed his weapon, helmet and field jacket. The subject unbottoned his fatigue shirt to expose his chest. A cotton alcohol swab was applied to the electrode emplacement site to remove any skin oil. An electrode paste reservior was prepared with electrode jelly and the electrode was emplaced on the subject.

The transmitter was then fixed to the subject's helmet with the aid of adhesive tape. The transmitter was connected to the electrodes with interconnecting leads which were passed under the shirt collar, down the subject's back and snap-fastened to the electrodes. Once the electrodes and transmitter were in place, the subject donned his equipment. The transmitter was turned on, (each transmitter had been tuned to an unused spot on the FM band), and the relaxation oscillator switched on to facilitate the tuning of the FM receiver. When the FM receiver was tuned, the transmitter was switched from the tuning position to the transmit position. Next, the receiver output was checked on the graphic recorder. After tuning was completed, the subjects returned to their section to begin an observational run. During their task activities, the variations in their heart rates, sensed and transmitted to the telemetry receivers in the nearby tent, were thus recorded as varying fluctuations in the lines being traced by the recorders along a time axis on rolls of graph paper. At the conclusion of the run, the three men returned to the telemetry tent to have the transmitting elements removed.

Difficulties - During the course of this study, it was found that the adhesive surface area of the electrode was insufficient to maintain good contact during severe motions of the subject. Consequently, additional adhesive tape was placed over the electrode to insure a high conductivity connection during periods of heavy exertion, i.e., when lifting the trails of the gun, running out aiming posts, or jumping off the truck.

The preparation time required to equip three subjects with the telemetry gear was approximately 30 minutes, which caused a delay between observational runs. This time was required, however, to adequately

prepare subjects, check the transmission, reception, and the graphic recording for each system.

The stability of the telemetry system was considered good. However, as the temperature and time of an observational run increased, considerable transmitter and receiver frequency drifts occurred. This frequency drift resulted in the degradation of the heart rate output signal displayed by the graphic recorder. In most instances, this degradation was corrected by the telemetry observer by: 1) changing the four volt Mercury cell battery in the transmitter between runs which eliminated loss of battery life, and 2) constantly monitoring the telemetry system to compensate for frequency drift. Generally speaking, the heart rate data recorded by this system was excellent.

h. Supplementary Film Recording

Implementation - The visual recording of additional orientation and procedural detail was with the following equipment:

- Super 8 mm Motion Picture Camera
- Twelve Super 8 Film Cartridges
- 35 mm Still Camera, Color Film

Procedure - This supplementary recording was done informally throughout the week of exercises by one of the film cameramen and the Contract Project Officer. The principal object was to document, for the project itself, those activities not officially covered: preparations by the technicians, briefing prior to the exercises, official visitors to the site, details on donning of masks and hoods, etc.

In between test periods, the film cameramen took "candid" 35 mm color pictures, and the Contract Project Officer completed several sequences, with Super 8 film, of the on-going activities. These proved quite useful in later briefing about field activity of analysis personnel not actually present on-site and in project presentations.

Difficulties - The lack of priority for this supplementary filming resulted in a lower level of presentational "polish" in some of the visual sequences. Many of the pictures duplicated those recorded by the other media, and this additional data collection was not really required. It was convenient to film on Super 8 and on 35 mm however, and the resulting color slides, prints and impromptu motion pictures added a perspective that aided post-field analysis and interpretation of the exercise. Magazine loading and automatic lens setting were particularly

helpful, for this candid pictorial recording, in overcoming the inconvenience of the field environment.

i. After-Task Interviews (Individual Subjective Reports)

Implementation - Since these interviews were conducted primarily through use of a simple questionnaire, the only required supplies were about 100 copies each of the two formats of the questionnaire: "With Mask" and "Without Mask," and a dozen pencils.

Procedure - The Oral-Visual Observer, after observing a given team, gave its members a standard briefing on the subjective purpose of the questionnaire, distributed pencils and the appropriate format, and answered any questions that arose during the filling-out of the sheets. (Results are reported in detail in Section IV).

Difficulties - All men were administered the questionnaire under both conditions and no major difficulties arose. Filling out the forms in the field, with no shelter nor writing surface, could have been most awkward if the forms had required extensive written comment. However, this trouble had been anticipated, and responses were limited to checks, single words and short phrases that were accomplished on the surface of truck fenders, a team member's back, etc. The immediacy of subjective impressions from the field task was thus preserved, but rainy weather would have demanded a minimal shelter for even this brief post-testing. Later analysis also revealed that a very few individuals had either failed to understand isolated questions or had deliberately falsified an answer. Since the Observer checked each form for completion as it was handed back, such data loss was negligible.

D. Summary of Field Exercise

The exercise served well its prime purpose of generating a set of field data for reduction and analysis in terms of the tentative design of a team test of protective equipment. The steps and results of the subsequent handling of this data are given in the following section and the implications, derived largely from this field experience, for standards of team performance evaluation are discussed in Section V.

In addition, the field exercise provided the participating battery some well-disciplined training under masked conditions, and the observation team with an insight into the practical problems in field team evaluation and the level of observer training required. Finally, the exercise was an actual field test of the various data gathering techniques, themselves. The performance of all data recording equipment and the field practicalities of the techniques are compared in depth in Section V.

Achievement of a field exercise yielding performance data, equipment evaluation and technique assessment--plus troop training values--was made possible both through careful preparatory design and through smooth cooperative efforts among Edgewood Arsenal Research Labs, its study contractor Dunlap and Associates, USCONARC, and the 82nd Airborne Division at Ft. Bragg.

SECTION IV

DATA HANDLING

A. Comprehensive Method of Data Collection

The data base for this study was six 105 mm howitzer teams performing a complete fire mission with and without the M-17 protective mask during the field observations at Ft. Bragg.

From these field runs, the data was collected through the following media, which correspond to the field technique described in Section III:

1. Visual and Oral Observer's check lists.

2. Oral Observer's and Team Umpire's personal tape recorder.

3. Video and audio recording on videotape.

4. Team Events Observer recording on one-track of the videotape.

5. Heart-rate of gunner, assistant gunner, and #3 man telemetered and recorded.

6. After-Task Questionnaires submitted by the team members.

7. Film camera recording (16 mm).

All these techniques to gather data from the field site were used to build the final data base. In many cases, the results of several techniques were integrated to form the complete picture of a variable. The multi-media approach used here actually contains many views obtained with the different measurement techniques; these were integrated to form a single comprehensive view of the pattern of field performance of the set of six teams.

The telemetry producing the heart-rate data, and the After-Task Questionnaires, were relatively independent measures and were considered directly from the media that produced them. The recording from the telemetry was the only source of heart-rate data. Also, the After-Task Questionnaires were the only source of the team members' opinions concerning their performance and certain aspects of the mask.

Other data, however, had several sources. Individual errors, for example, were derived from the Team Events Observer's recording, the videotape, and the films. In this case, each source of data complemented

the other, resulting finally in the extraction of all personal errors made in the performance of the task by that team with and without the mask.

To obtain data from the videotapes, modified check lists and Data Sheets were made after the observers returned from the field study. Since the actual field task was slightly different from the original format used by the observers on site at Ft. Bragg, the new formats were tailored to review of the videotapes of the team's performance. This allowed the observers to follow the events more closely and obtain their data more completely in a schedule of three runs with a minimum of stopping and backtracking of the tape. These revised versions of the task outline thus reflected the task specifically as it was conducted at Ft. Bragg. Other Data Sheets were constructed to record errors, quality, and times. The times were extracted by using a series of four observers with stopwatches monitoring specific times of the individual men and the sub-tasks.

1. Definition of Data Terminology

Videotaping (including the audio tracks on the videotapes) played a predominant role in effective data collection for this study, with the other methods supplementary. Integrating all of the data sources as mentioned, the categories of data were then defined in the following terms:

"Orals: Scheduled and Unscheduled"

Oral Intercommunication among team members included spoken signals from one man to another or others. Certain of these signals are necessary for successful completion of the task. For example, the deflection command from the Fire Direction Center must be relayed to the gunner in order that the Howitzer be fired in the proper direction. Oral signals of this nature were termed and scored as "scheduled orals." Other verbal interchange, not basically necessary according to standard operating procedures, but which did take place were termed and scored as "unscheduled orals." Any oral signal that was repeated was called an "oral repeat." The oral intercommunications were best obtained from the soundtrack of the videotape. Several runs of the videotape proved this to be a more exhaustive and reliable source of oral intercommunication than the observer's check list used in the field. There were, however, some oral signals that the observer in the field could check and which could not be heard on the soundtrack. In this case, the check list and the video soundtrack were compared to detect the additional data. The observer's notes on communications from his personal audio recorder were transcribed and also integrated into the total count and analysis of oral interactivity.

"Visuals: Scheduled and Unscheduled"

The visual intercommunication measure was a record of all visual

signals used to convey information from one team member to another. As in the area of oral signals, certain visual signals are required for task success. The gunner directs the #4 man in placing the aiming posts by visual signaling according to standard operating procedures in the Howitzer task. This would be counted as a "scheduled visual." Others, such as the Chief-of-Section signaling one of the men to move faster, is not normally a required procedure, although it contributes to the specific performance. Signals of this nature were counted as "unscheduled visuals." Any signals that were repeated were counted as "visual repeats." In this type of measure also, the videotape proved to be a reliable source. The videorecorders covered the total Howitzer site with a long-range, or total-team, camera and a closeup camera. A visual signal observed at a distance, via the total-team camera, could often be reviewed on the closeup videotape for definite classification as scheduled, unscheduled or repeat.

Errors: "Personal" and "Team"

An error count was compiled according to three categories. Any discontinuity in behavior of an individual team member due to accident or physical shortcoming, such as stumbling over equipment, running into another man, or dropping equipment, was judged as a "personal error." Other errors, committed by the team as a group or by an individual through lack of team training or coordination, were called "team errors." These would be behaviors such as two or three men moving the trails while the gunner was trying to lay the bore in the proper direction.

"Unsafe Conditions"

These were officially determined by a safety officer participating in the exercise, and unofficially by the Team Events Observer, who was an experienced battery officer. Another role of this observer was to determine "critical incidents" that happened during the task. These were generally unsafe team conditions that existed or unsafe behaviors of the team members not necessarily resulting in accidents in the specific instance. For example, a man standing in the recoil area of the Howitzer after the shell was loaded would be tabulated under critical incidents, even though his unsafe behavior had no serious consequence on this occasion. The measure of errors was obtained mainly from the videotape, and was supported by data collected from the Team Events Observer's recorded soundtrack and the 16 mm color films.

Time Data

The time data was taken from the videotape. The running time of each sub-task was part of the videotape. A digital time counter was inserted in the upper corner of the visual display. Other times were obtained by

observers using stopwatches while monitoring running sub-task sequences on the videotapes.

Quality Data

The quality of team performance was subjectively judged by the Team Events Observer as an experienced battery officer on Howitzers. He rated each sub-task in the field as "good," "fair," or "poor," on the check list and on the videotape's soundtrack. He also rerated the sub-tasks later when viewing the videotapes of the teams participating. On many of the tasks, another observer also rated quality as a consistency check on this subjective rating technique.

2. Definition of Data Categories

a. Base Data

Section III of this report explained the overall operations used to gather data from the observer's field recordings and check lists, and the post-field gathering of more specific data from the audio and videotapes. The data collected in these several modes was integrated, for reduction and analysis, into the following basic categories of variables:

Oral Intercommunication

(1) Scheduled orals per sub-task

(2) Unscheduled orals per sub-task

(3) Oral repeats per sub-task

Visual Intercommunication

(1) Scheduled visuals per sub-task

(2) Unscheduled visuals per sub-task

(3) Visual repeats per sub-task

Error Count

(1) Personal errors per sub-task

(2) Team errors per sub-task

(3) Critical incidents per sub-task

Time

(1) Total time per sub-task

(2) Working time for each man per sub-task

(3) Aiming-post time for sub-task two

Quality Rating

(1) Quality per sub-task

Heart-rate

(1) Monitored heart beats for the gunner, assistant gunner, and #3 man throughout all sub-tasks.

After-Task Questionnaires

(1) Rating of team and individual proficiency by team members at the end of each whole task.

(2) Check list concerning attitude toward

- fit of mask
- vision with mask
- speaking ability with mask
- adaptation to the mask
- feeling of tiredness with mask
- benefit of further practice with mask

(3) Biographical data

- Name, rank, MOS, and team position
- Time in service and in training, both on Howitzers and with present team members
- Number of fire missions completed
- Understanding of the purpose of the study

As previously noted, the Howitzer task was divided into five sub-tasks or phases of activity, making up the total task. The base data is predominantly in comparative sub-task form, except for the particular types of measures which could be gotten only on the whole task basis. The After-Task Questionnaire is an example of this case. Also, the time to place aiming-posts is an isolated sub-task measure and was extracted from sub-task two.

b. Combinations of Data

Certain measures obtained, although deemed important, occur rather sparsely on a sub-task basis (critical incidents, e.g.) and are, therefore, considered only in the sub-task where they predominently occur, or are added for all five sub-tasks and considered on a whole task basis.

In other cases, combinations were made in order to form a comprehensive category which was expected to yield a discriminative effect between the mask and no mask condition. These combined measures were generally viewed on a whole task basis, and a sub-task analysis was done for the dominant sub-task producing the effect.

These combined measures are defined as follows and apply to the task or sub-task approach:

1. Combined data on a task basis:

   Total = Sub-task 1 + . . . + Sub-task 5

   This is the sum of all sub-tasks for the variables mentioned in the list of base data, except for heart-rate, the After-Task Questionnaires and quality. Quality on a task basis was averaged due to the inherent cumulative nature of this measure; quality cannot be assessed from isolated events.

2. Total orals = Scheduled orals + Unscheduled orals + Oral repeats

3. Total visuals = Scheduled visuals + Unscheduled visuals + Visual repeats

4. Total communication = Total visuals + Total orals

5. Total errors = Personal errors + Team errors + Critical incidents

6. Communications deviations = Unscheduled visuals + Visual repeats + Unscheduled orals + Oral repeats

7. Total scheduled communication = Scheduled visuals + Scheduled orals

8. Total deviations = Total errors + Communication deviations

9. Density = $\frac{\text{Working Man Time}}{\text{Total Time}}$

B. Reduction and Analysis of Data

The approach used to view the obtained data was governed by the need to determine which variables, or combination of variables, could effectively discriminate between the protective mask and no protective mask conditions. Keeping this basic criterion in mind, there was the further consideration of developing recommendations for a structured test, including many of the group variables in this task chosen as an experimental data base. To do this, the analysis of the pattern of collected data must include an investigation of such questions as:

1. What is the variable, or combination of variables, producing a discrimination between the task performed with and without the mask?

2. What is the status of the measured variable; what is its magnitude, form of distribution, and reliability?

3. Is the discrimination significant statistically? What factors in the experiment, favorably or unfavorably, act to alter the results?

4. Is the effect best shown on a sub-task basis or whole task basis; if sub-task, which sub-task is most relevant for the given variable?

5. How does the variable under consideration relate to other variables in the experiment?

As implied in the outline of the sub-tasks in Section II, certain sub-tasks were expected to produce greater magnitudes of some variables. For example, sub-task 2, laying of the Howitzer, should have the greatest amount of visual signaling. This could be seen from the task analysis and proved true in the experiment. Other variables, team errors for instance, proved to be about equally distributed throughout the five sub-tasks. The analysis, as a result, was done in two phases. The first phase of analysis was done on a gross whole task basis, and a second phase investigated the significance of the particular sub-tasks. Other effects, which became apparent either from the data itself or from the effect of the experimental design on the data, were also considered.

1. Gross Analysis on Whole Task Basis

a. Mask Effects

The whole task basis allows a view of the data for the complete tactical mission of the Howitzer team. Several types of analyses were employed here. Shown in Table A are correlated t's done on the whole task for six teams participating. The means and s.d.'s with and without the mask are given as well as the t ratios. The stars indicate a significant effect, and the dashes indicate that no test can be performed due to no distribution or to lack of homogeneity of variance. It can be seen that two of the combined measures are significant. These are the total visual signals employed and the total communication employed. The total visual count is itself a component of total communication; but the advantage of the combined measure is that it often presents more of a variable to work with. It often is homogeneous even when its independent components may be sparse (unreliable measures) and not homogeneous individually. The combination is intended to be a more comprehensive general group trait as well. "Total errors," for instance, is a combination of team errors, personal errors, and critical incidents. This yields a more general trait of errors as a team variable. Most of the data in Table A exhibits the expected direction, but is not significant on the whole task basis.

b. Regression Analysis

By regressing from a combined measure to its separate components through a statistical device that assesses their relative contributions, the most important variables in producing the total measure can be singled out. Accordingly, a regression analysis was made by running a series of multiple R's on the SDS-920 computer at Dunlap and Associates. The resulting figures, which indicate the "weight" of each variable upon the combined measures, are given in Table B. (For total orals and total communication, one or more of the components are missing so that the resulting multiple R's are lower). For the total oral interactivity measure, the unscheduled orals contribute the most variance, and this is true for both the without and with mask conditions as can be seen from greater magnitudes of the figures in both boxes under the "unscheduled" column. For the visual signaling interactivity, the scheduled visuals contribute most to the variance of the total. On total errors, team errors weight highest in the without condition, but team errors and personal errors weight equally in the with condition. Concerning the total communication, "unscheduled orals" is the dominant factor contributing to the total, with and without the mask. (The weightings shown in Table B are the derived β^2 weights and are best viewed as those measures which most strongly effect the prediction of the total score, given that combination of components in the make-up of the total score). A knowledge of which variables are most operative in the data base will allow a more sensitive test structure, as discussed further at the end of Section V.

Table A

Overall Results of Tests of Significance Between Mask and No Mask Condition Based on the Whole Task

Variable	M w/o	M w	s.d. w/o	s.d. w	t corr.
Sched. Verbal	46.5	42.8	4.42	4.62	1.80
Unsched. Verbal	12.3	16.8	8.38	9.89	--
Verbal Repeats	0	1.3	0	1.37	--
Total Verbals	59.2	60.3	10.28	8.73	0.88
Sched. Manual	25.5	32.3	6.02	7.34	1.50
Unsched. Manual	2.8	8.7	2.32	5.35	2.06
Manual Repeats	0.6	2.8	0.84	2.07	--
Total Manuals	28.8	43.3	5.98	10.39	3.00*
Personal Errors	2.8	5.2	2.86	2.64	1.59
Team Errors	5.0	5.2	2.10	2.64	0.83
Critical Incidents	0.50	0.83	0.55	0.98	--
Total Errors	7.8	10.3	3.71	3.61	1.90
Quality	2.4	2.3	0.38	0.51	0.16
Total Time	9.6 m	11.0 m	1.62 m	3.10 m	1.25
Aiming Post Time	56 sec.	61. sec.	21.2 sec.	17.1 sec.	0.46
Total Commun.	88.0	103.7	10.90	14.43	2.60*
Total Sched. Com.	70.6	75.2	4.4	9.3	1.04
Total Unsched. "	16.0	30.3	8.1	15.2	2.11
Total Deviations	24.6	41.5	7.0	16.2	2.36

N = 6 teams w/o = without mask

$t_{crit.\ .05} = 2.57$ w = with mask

Table B

Multiple Regression Analysis for the Combination Measures on a Whole Task Basis

1. Total Oral

	Sched. Oral	Unsched. Oral	Oral Repeats
Without	.080	0.50	0.00
With	0.22	1.24	.004

2. Total Visual

	Sched. Visual	Unsched. Visual	Visual Repeats
Without	1.01	0.15	0.02
With	0.49	0.26	0.04

3. Total Errors

	Team E	Personal E	Critical Incid.
Without	0.60	0.32	0.0
With	0.53	0.53	0.0

4. Total Comm.

	Sched. Oral	Unsched. Oral	Sched. Visual	Unsched. Visual
Without	0.0006	0.60	0.35	0.00
With	.032	1.13	0.01	0.22
Combined	0.15	0.82	0.07	0.15

Magnitude of "weights" indicate ability to predict total score. The higher numbers will better predict (have greater effect) on the total.

c. Sequence Effects

One noticeable outcome of the whole task results was an interaction between mask effects and the order of run of the two conditions by the same team. It was expected that, with inexperienced teams, some general learning of the task would take place regardless of mask condition, and a counterbalanced design was used to attempt to cancel out this learning effect. That is, if team 1 performed the task without a mask and then performed the second time with a mask, team 2 would perform the task with a mask first and then without a mask. This format would be followed for the rest of the teams. In general, the design was effective, but in many cases the apparent learning effects were very large or very different between sequences, possibly because learning to adapt to the mask was unavoidably involved. Mask effects then became comparatively small, and the statistical outcome was affected. When sequence effects are large, the distribution widens and the mask effect is made to appear statistically insignificant. In order to demonstrate the true relative effects due to order of performance and protective masks, a profile presentation was drawn for each of the variables outlining sequence of performance versus mask and no mask conditions of the respective teams. Two simple formulas were derived from a hypothetical profile displaying mask effects (differences between times under like condition) versus sequence effects (general slope of lines from average first condition to average second). A brief explanation follows.

If an ideal example is assumed in which all teams have reached a plateau of efficiency prior to either trial, no learning effect will take place from first to second performance. Any slope in the line displayed between a first and second trial should then be attributable to the mask effect, not learning. Figure 3 shows this hypothetical profile along a graph with time and sequence coordinates. The example team performs first without mask in ten minutes, then with mask in twenty minutes. Its counterbalancing team performs first with mask in twenty minutes, then without mask in ten minutes. The obvious effect of the mask is to increase performance time by ten minutes. Next, an example is graphed which incorporates the more realistic assumption of learning from first to second trial in the sequence. Figure 4 shows the "without mask first" team still taking ten minutes initially, but taking only fifteen minutes on their second run, while wearing masks. The "mask first" team still takes the full twenty minutes initially, but now takes only five minutes on the second run when doubly benefitting from learning and lack of mask encumbrance. Both teams have cut five minutes from their performance on second attempt, due to an overall learning factor. This is demonstrated in the figure by the equal downward displace-

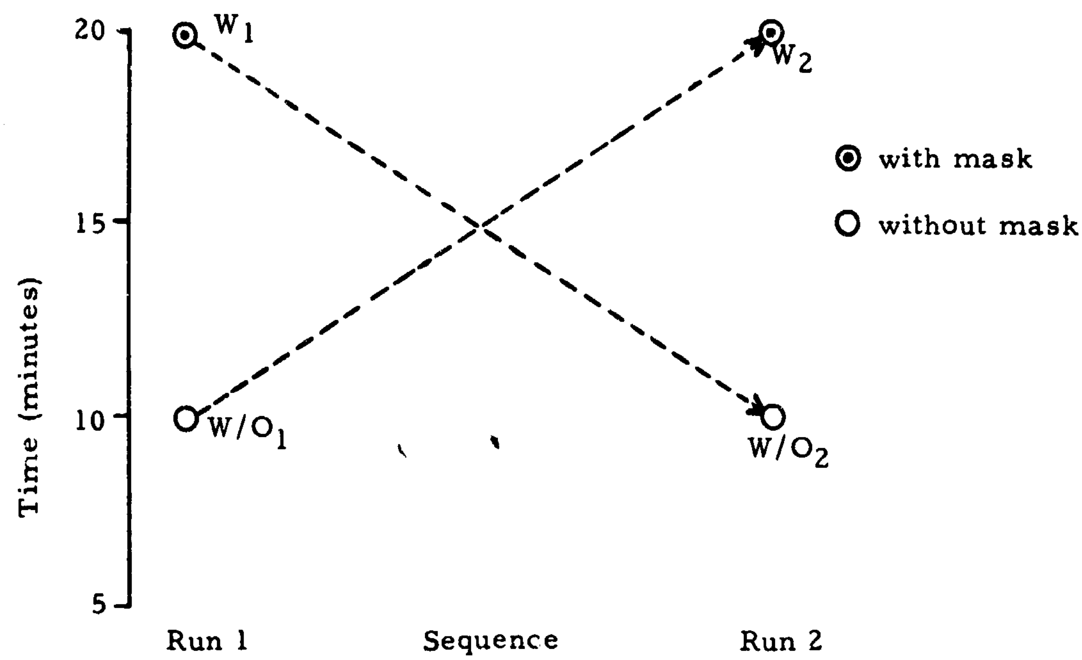


Figure 3 . Profile lines of ideal counterbalanced design showing mask effects and no sequence effects.

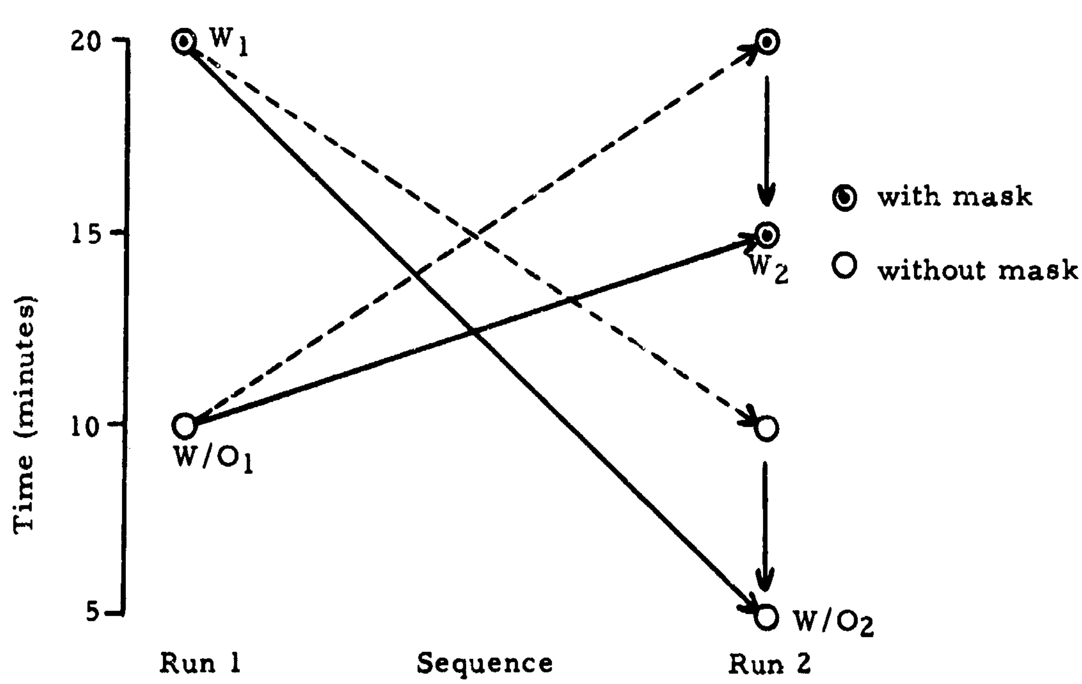


Figure 4 . Comparison profile lines of a counterbalanced design showing the ideal (dashed lines) and a profile with mask effects and sequence effects (full lines)

ments of the two second trial points. * The same ten minute mask effect still remains when times for both teams are averaged (17.5 - 7.5 = 10), but the differing absolute slopes of the two sequence lines, caused by introduction of the learning effect, now obscures this symmetry, so apparent in the ideal profile. However, a formula can now be derived from the graphic model to reveal the average sequence effect. This formula will average the differences between the initial reading of a variable and its obtained value on the second run: (With Mask_1 - Without Mask_2), divided by 2. Arranging this so that a positive answer indicates an increase in the variable considered from run 1 to run 2, we have:

$$\text{Avg. Seq. Effect} = \frac{(W_2 - W_1) + (W/O_2 - W/O_1)}{2}$$

W : with mask
W/O: without mask
1 : first run
2 : 2nd run

The mask effects can be calculated by simply averaging the mask effects and no masks effects and noting the difference.

$$\text{Avg. Mask Effect} = \frac{W_2 + W_1}{2} - \frac{W/O_1 + W/O_2}{2}$$

In this case, a positive value indicates an increase with the mask on. It can be noted concerning the slope describing the average sequence effect, that the greater the slope the larger will be the distribution of the variable, possibly approaching a bimodal distribution. This makes a significant result for the experimental variable very difficult to achieve. Table C outlines the average mask and sequence effects. The mask effects are no different than the difference between means on a task basis, but can now be compared with the average sequence effect applicable to that variable. Each point on the profile line presentation is the average of those teams performing under that condition--that is, run one with a mask, run one without a mask, run two with a mask, run two without a mask. This breakdown, shown graphically by lines, makes more clear the effects of these variables in the field exercise. The general result from Table C is that the mask had the effect of increasing communications, errors, time, and deviations. In the case of quality and scheduled verbals, the mask had a negative effect. The effects due to sequence, on the other hand, are generally negative; a decrease is seen in communications, errors, time, and deviations. Quality

---

* We will ignore for our present purpose the even more realistic fact that the nature and possibly the magnitude of the learning effect would actually be different for the team attempting to adapt to the mask while performing for the first time, than for the team learning to adapt to the mask on its second trial, after having already learned the task without the mask.

Table C

Effects of Mask and Sequence of Performance
on the Whole Task Basis

Variable	Mask Effect Data	Mask Effect %	Sequence Effect (2nd Run) Data	Sequence Effect (2nd Run) %
Sched. Orals	- 1.8	- 7.7%	2.0	4.5%
Unsched. Orals	4.5	36.0%	- 1.8	- 11.5%
Oral Repeats	1.3	--	- 1.0	- 87.0%
Total Orals	1.5	2.0%	- 0.5	- 0.8%
Sched. Visuals	3.7	13.0%	- 7.8	- 22.5%
Unsched. Visuals	5.9	110.0%	- 3.9	- 51.0%
Visual Repeats	1.9	310.0%	- 0.5	- 28.0%
Total Visuals	15.5*	54.0%	- 9.0	- 22.5%
Total Commun.	15.7*	18.0%	- 9.3	- 9.2%
Personal Errors	2.4	86.0%	- 1.0	- 22.5%
Team Errors	0.2	4.0%	- 1.9	- 32.0%
Critical Incidents	0.0	--	0.0	--
Total Errors	2.8	34.0%	- 2.5	- 22.5%
Quality	- 0.2	- 9.0%	0.1	30.0%
Total Task Time	1.4 min.	14.5%	- 2.3 min. *	- 20.0%
Aiming Post Time	6.5 sec.	12.0%	- 24.0 sec. *	- 34.0%
Total Unsched. Comm.	14.4	90.0%	- 9.3	- 33.5%
Total Sched. Comm.	3.0	4.0%	- 2.4	- 3.2%
Total Deviations	16.9	69.0%	- 12.3	- 32.0%

* sig. at .05 level

and scheduled verbals increase on the second performance. In many cases, the sequence effect far exceeds the mask effect. Aiming-post time is an example of this. As shown in Table C, the mask adds an additional 6.5 seconds on the average, but on the second run it is decreased by 24.0 seconds on the average. An example of the use of the profile presentation to detect and demonstrate this predominance of sequence effect is shown in Figure 5. The general slope of both lines, whether sequence is from the average of the teams who went with or without mask first, is quite steep. The variables other than times in this study are plotted in the same manner and shown in the appendix of the report. The described method of graphic presentation is used for all as a practical means of displaying these interacting factors.

The factor other than desired mask effect has been called sequence, or order effect, since it makes itself evident on the second trial. It may be inferred that it is some type of learning that has taken place on the first run of the task. Whether it is that the team members learn the task, how to work together, or the experimental setting, is difficult to say. Since the task is run the second time in the same emplacement position, and in this case the aiming posts are put in about the same place, we may infer a learning of the "process in that setting" as one of the strongest contributors to the sequence effects. A point worth noting here is that observer adaptation to monitoring a given team the second time around is probably not a serious contributing factor to the sequence effect, since most data was collected, not on-site in the actual field sequence, but from the videotapes in viewings not necessarily in the original order of performance.

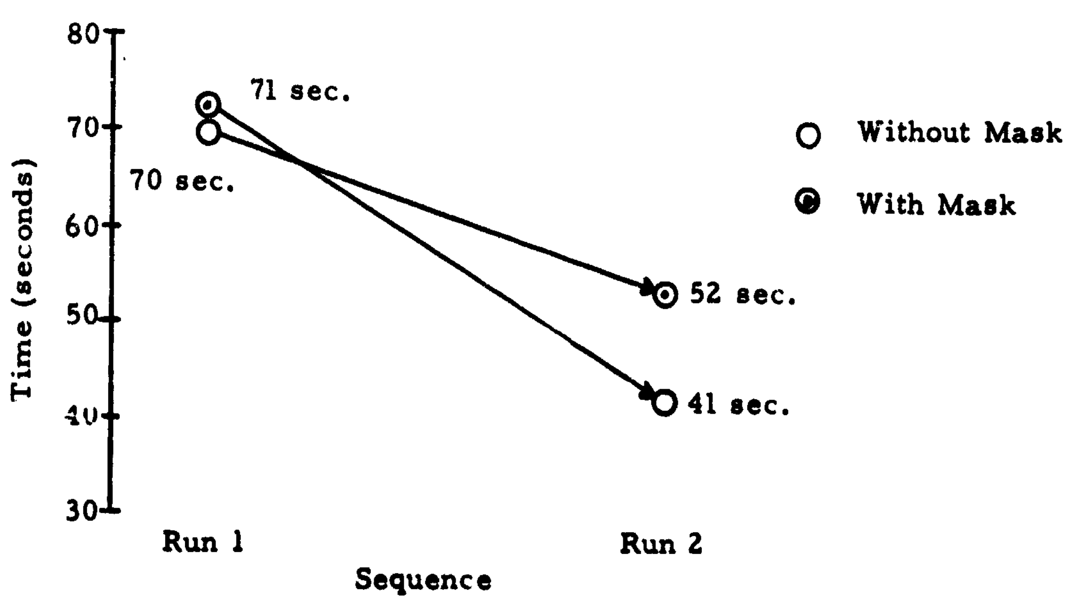


Figure 5. Profile Analysis of Mask and Sequence on Aiming Post Time

A 2 x 2 analysis of variance was calculated for the sequence (1st and 2nd) effects and mask (with and without) effects. The significant differences were the same as those found in the table for the correlated t's (see Table A). Significant sequence differences are total time and aiming-post time. These are shown in Table C.

d. <u>After-Task Questionnaires</u>

The After-Task Questionnaires used in the study were described in Section III. The results from that questionnaire are reported here.

Six teams, each comprising 8 men, filled out the report after performing the task with and without masks. Thus, 48 subjects filled out the report twice, once under each condition.

(1) Experimental Population

Biographical data describing the population sufficiently for this study was obtained from the After-Task Report rather than from more detailed personnel files. Of the total individuals making up all subject teams--all chiefs of section (C/S), gunners (G), assistant gunners (AG), and cannoneer positions Nos. 1 to 5--the rank of the C/S's was either E-5 or E-6 and the rest of the men had a rank of E-2 or E-3. (Eighty-five per cent were E-2's and the rest E-3's, and all were in active training).

All of the participants had artillery MOS's. Most of the men (75%) were in their normal team positions during the experiment. Those who were not had very slight changes of duty slot; for example, a number 2 man may have taken a number 3 position.

The C/S's exhibited a range of 10 to 18 years in the service, while none of the men in their sections were in longer than 13 months. The C/S's average about 10 years on howitzers, and the men ranged from 2 to 6 months. Concerning the time a team had worked together on the task, the C/S's ranged from 1 day to 1 month, and the men 1 day to 2 months.

The men's history of experience with the mask showed that 75% had worn masks before and 42% had worn masks while doing some form of work on a howitzer task, but not more than a few times on the average.

Since the study was done over several days, some men happened to be team members more than once. Of the total, 9 men participated a second time and in the same position on both teams.

(2) Attitude Toward Mask

The After-Task Questionnaire was designed primarily to

gather information helpful for the study, but not objectively available to the observers; information that is related to the subject's feelings toward the mask while performing his tactical task. Table D shows the results based on the total N of 48 completed questionnaires. In general, there was no outright condemnation of the mask by any respondent. The only strong negative response seemed to be to the question relating to tiredness.

(3) Attitude Toward Performance

Other than the biographical data and attitude toward the mask, other questions relative to the men's performance were included for both mask and no mask formats. This made feasible a comparison of an individual's subjective ratings for his with and without mask conditions. The men were also asked whether their team's performance was "better," "as well," or "worse," than "normal;" and whether their own performance was "better," "as well," or "worse," than "normal." The results are shown in Table E. These are shown as percentages based on the N of 48 responses. The table demonstrates an increase in the percentage of individual men answering "worse" when in the mask condition; the greatest difference being an increase in "worse" responses on individual performance ratings from 15% without mask to 31% with mask. A chi square test of independence (3 x 2) yielded no statistically significant difference, however, between the without and with mask groups in their answers to all three categories.

A series of correlations was run between responses of "worse," "as well," and "better," and the other variables obtained. This was done on a whole task basis, for the with and without mask condition, and for the responses of attitude toward individual and team performance, and included intercorrelations between attitude on individual and team performance. The results are shown in Table F (the stars indicate correlations significant beyond the .05 level of probability). From the chart it can be seen that those men answering "worse" in rating own team performance also answered "worse" in rating their own individual performance, holding true for both with and without mask conditions. Two effects may be generally noted on the charts. A "better" or "worse" response given for team performance anticipated a "better" or "worse" rating for that individual's own performance. This may be a valid generalized averaging of individuals' feelings, or it may be a sequenced response bias in order to finish the After-Task Questionnaire more quickly. It can also be noted that the significant effect in response to the median attitude, "as well," exists for the without mask condition, but not for the with mask condition. With the mask, however, the two extremes of response, "better" and "worse" seem more definitely correlated.

These intercorrelations are not independent by virtue of

Table D

Responses to Questions on Attitude Toward the Mask

Question	Yes	No	No Ans.
Do you view the mask as a lifesaving device?	98%	2%	--
Was the fit comfortable?	96%	4%	--
Could you see well enough to do your job?	85%	10%	5%
Could you speak easily to the others in team?	85%	15%	4%
Did you feel more tired with the mask on?	50%	48%	2%
Did you feel you were getting used to the mask?	81%	19%	--
Would practice increase your ability to work with the mask on?	81%	19%	--
Was the purpose of the study made clear to you?	83%	5%	12%

Table E

Responses to Questions on Attitude Toward Own Performance

Team Performance:

	Without Mask	With Mask
Better	15%	14%
As Well	58%	50%
Worse	21%	35%
No Ans.	6%	2%

Individual Performance:

	Without Mask	With Mask
Better	19%	11%
As Well	65%	56%
Worse	15%	31%
No Ans.	2%	2%

Table F

Intercorrelations Among Responses Toward
Individual and Team Proficiency

	With Mask	Better	As Well	Worse	Team
Individual	Better	0.920*	.409	- .812*	
	As Well	0.570	0.570	-0.781*	
	Worse	-0.831*	-0.511	0.897*	

	Without Mask	Better	As Well	Worse	Team
Individual	Better	0.739	-0.310	0.000	
	As Well	-0.173	0.928*	-0.832*	
	Worse	-0.309	-0.881*	0.985*	

Table G

Intercorrelations Among Subjective and
External Proficiency Measures

	Without Mask	Total Errors	Quality	Time
Team	Better	0.739	0.082	0.494
	As Well	-0.006	0.796*	0.568
	Worse	-0.287	-0.732	-0.767*
Individual	Better	0.880*	-0.112	0.035
	As Well	-0.192	0.588	0.556
	Worse	-0.318	-0.665	-0.713

	With Mask	Total Errors	Quality	Time
Team	Better	-0.198	0.185	-0.350
	As Well	-0.893*	0.694	-0.884*
	Worse	0.677	-0.645	0.746
Individual	Better	-0.316	0.309	-0.454
	As Well	-0.422	0.665	-0.628
	Worse	0.384	-0.662	0.589

* Significant at .05 level

the construction of the forms: A man cannot answer "better" and "worse" at the same time. This explains the indicated negative correlations. The correlation between "worse" and "as well" is significantly negative for the with and without mask condition--the greater the "worse" responses, the less must be the "as well" responses given.

The relation between the men's personal attitudes toward their own and their team's performance with relevant external measures yields some further results of interest. Table G shows the correlations among these subjective measures and the objective measures of total errors (personal and team), quality rating, and time (all on the basis of the whole completed task). The comparison reveals no obvious relationship among the men's opinion of performance (theirs and their teams) and more objective measures which may relate to proficiency. Unless only the poorest teams felt subjective improvement during the exercise over their normal performance, it might be expected that individual ratings of "better" would correlate positively with the expert's objective quality ratings, but this was not the case. On the contrary, a negative but not significant correlation exists between number of "worse" responses and quality in all cases. There appears to be no definite relationship between subjective assessment of performance and external measures of that performance; self-ratings, at least in the form attempted by this study, offered no consistent basis for team performance evaluation.

e. Hoods Condition

In two of the test runs at Ft. Bragg, an M-6 hood was added to the M-17 mask condition. These test runs were conducted in the same manner as the ones with and without mask, except that no telemetry was included and no After-Task Questionnaires were administered.

As in the case of the whole task data, the results with the hood condition must be viewed in terms of the particular sequence of a team's performance. Only two teams performed with hoods and their sequences were not the same. Consequently, there is no exact replication of either of the two teams and no statistical testing can be done. The raw data can be viewed, however, in the order in which the teams performed under the three conditions in Fig. 6 & 7. In this case, each of the two teams performed three times. One team ran the sequence (1) with mask, (2) with mask plus hood, and (3) with no mask. The other team ran the sequence (1) without mask, (2) with mask, and (3) with mask plus hood. Unfortunately, there were no further opportunities to test with the hoods in the field.

The first profile shown in Figure 6 outlines the time to complete the total task under the three conditions. There is a decrease in time throughout all three runs of both teams regardless of the experimental condition,

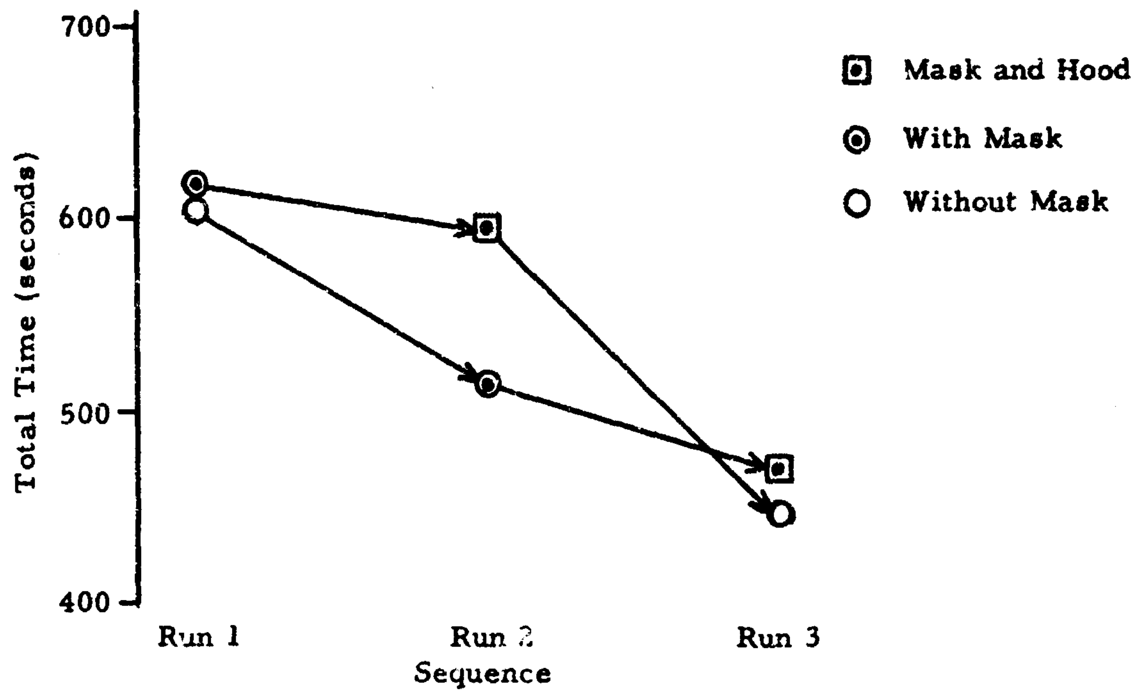


Figure 6 . Profile Lines of Total Time per Task

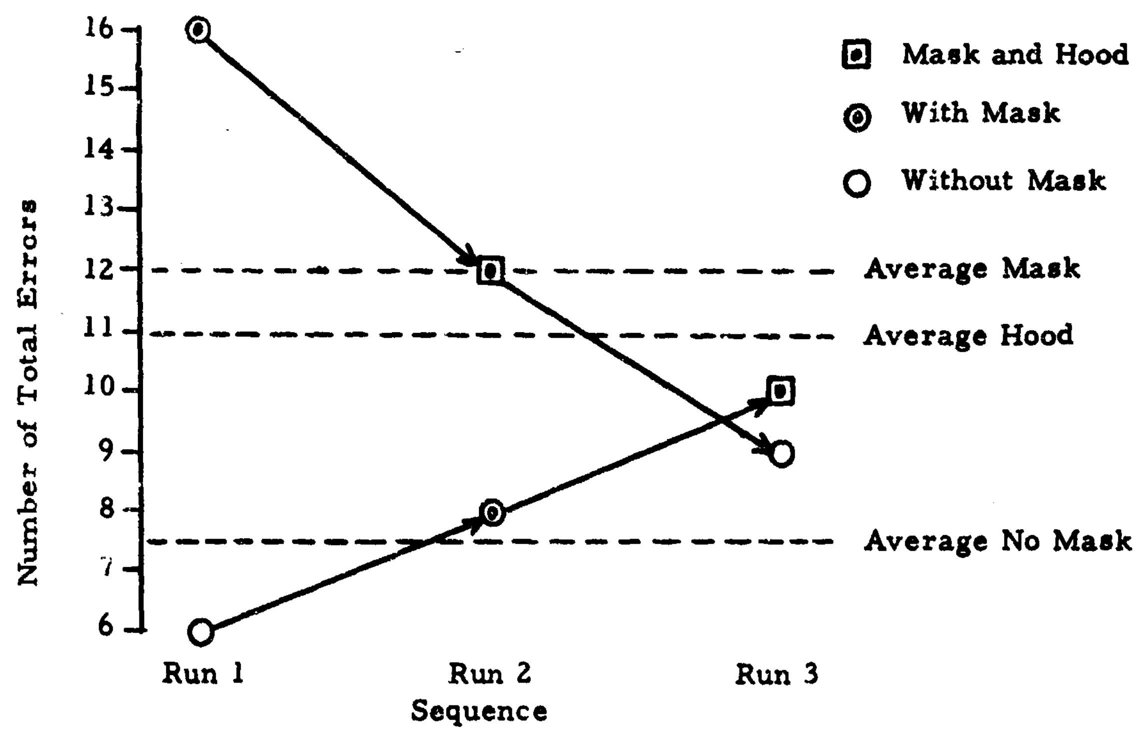


Figure 7 . Profile Lines of Total Errors per Task
(Team Errors, Personal Errors, Unsafes)

again demonstrating the dominant sequence effect. However, the gradient is much steeper when going from a mask or mask plus hood run to one without mask than when going from a mask to a mask plus hood condition. This suggests that the difference is very small between the conditions of mask alone and mask plus hood, and much larger between mask or hoods condition and no protective equipment at all.

On measures of team intercommunication (profiled as Supplementary Data items), one team showed an increase with the hoods for both oral and visual modes of signaling. The other team showed an increase with hoods for the oral, but not the visual communication. Once again, though, the sequence must be kept in mind. The mask first appeared to facilitate the without mask condition that followed. But "without the mask" first seldom provides learning helpful to the "with mask" condition that follows.

In general, an average score of both teams shows little difference between the mask and hoods condition, but a greater difference between either of these and the without mask condition. The profile on errors in Figure 7 shows this by designating the average errors for both teams by dashed level lines for each condition. Here the difference between the mask and mask plus hood is much less than the difference between either of these and the no mask condition.

This limited discussion must not be generalized, however, until more than two teams can be observed. Sequence effects are still strong regardless of using a third experimental condition.

2. Analysis on a Sub-task Basis

An investigation of the data from the howitzer task as a whole was the subject of the last several pages. It remains to view the component sub-tasks for possible predominant effects, within specific activities, of a variable which, averaged over the whole task, may have appeared unimpressive.

a. Sub-task Bar Graphs

The initial approach was to reduce the data for each variable to sub-task totals for both conditions in order to make use of a standard presentational structure for rough comparison, by visual inspection, among sub-tasks for the effects of the variables. Two simple five-fingered bar graphs were combined, for both conditions of a given variable, into a single comparison profile of the task in which five pairs of sub-task data for mask and no mask condition stand side by side. This presentation thus allows easy visual inspection of the data, both for the sub-task in which a variable has its greatest magnitude and for the sub-task in which there is maximum

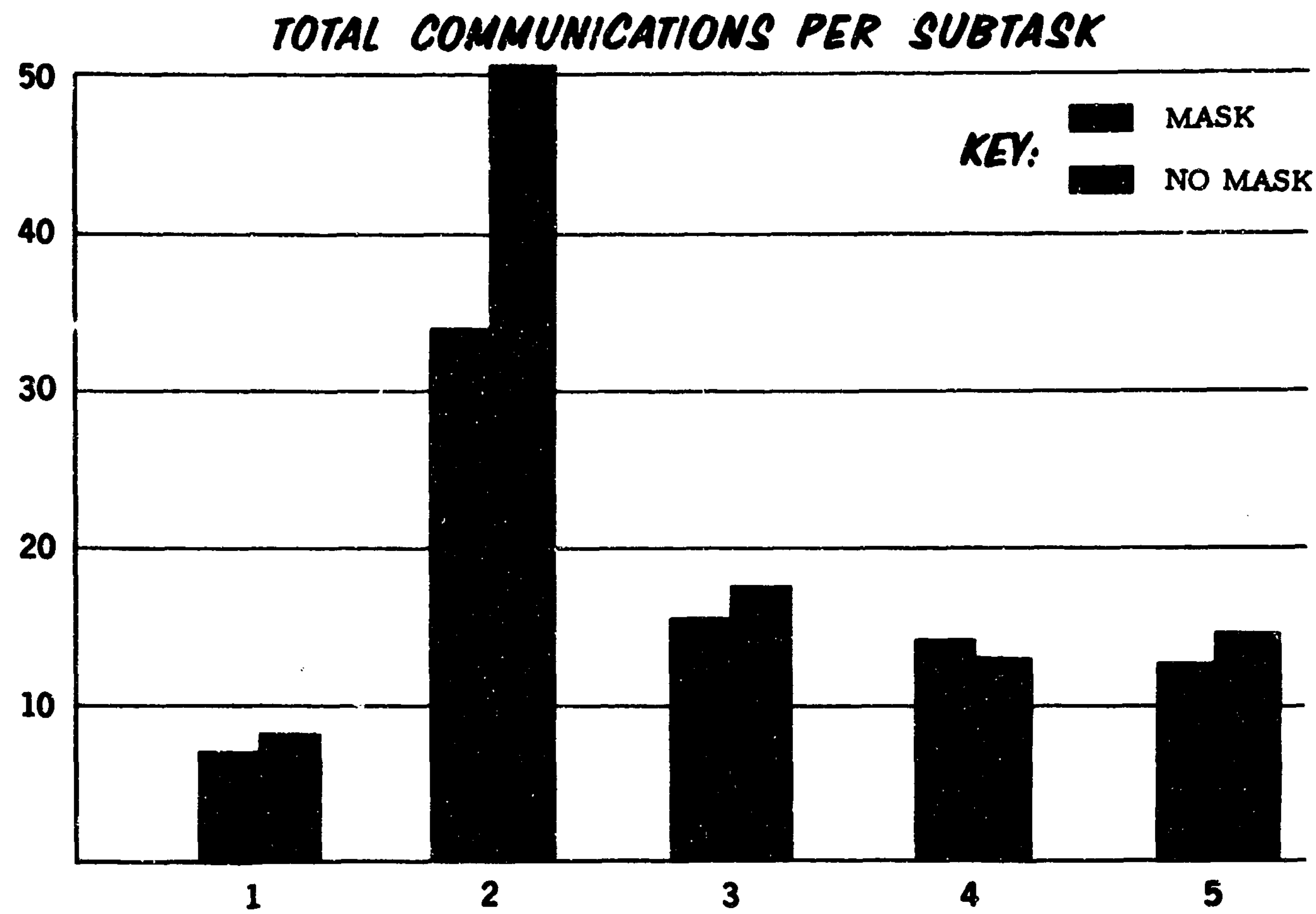


Figure 8. Bar graph of average number of total communications per subtask.

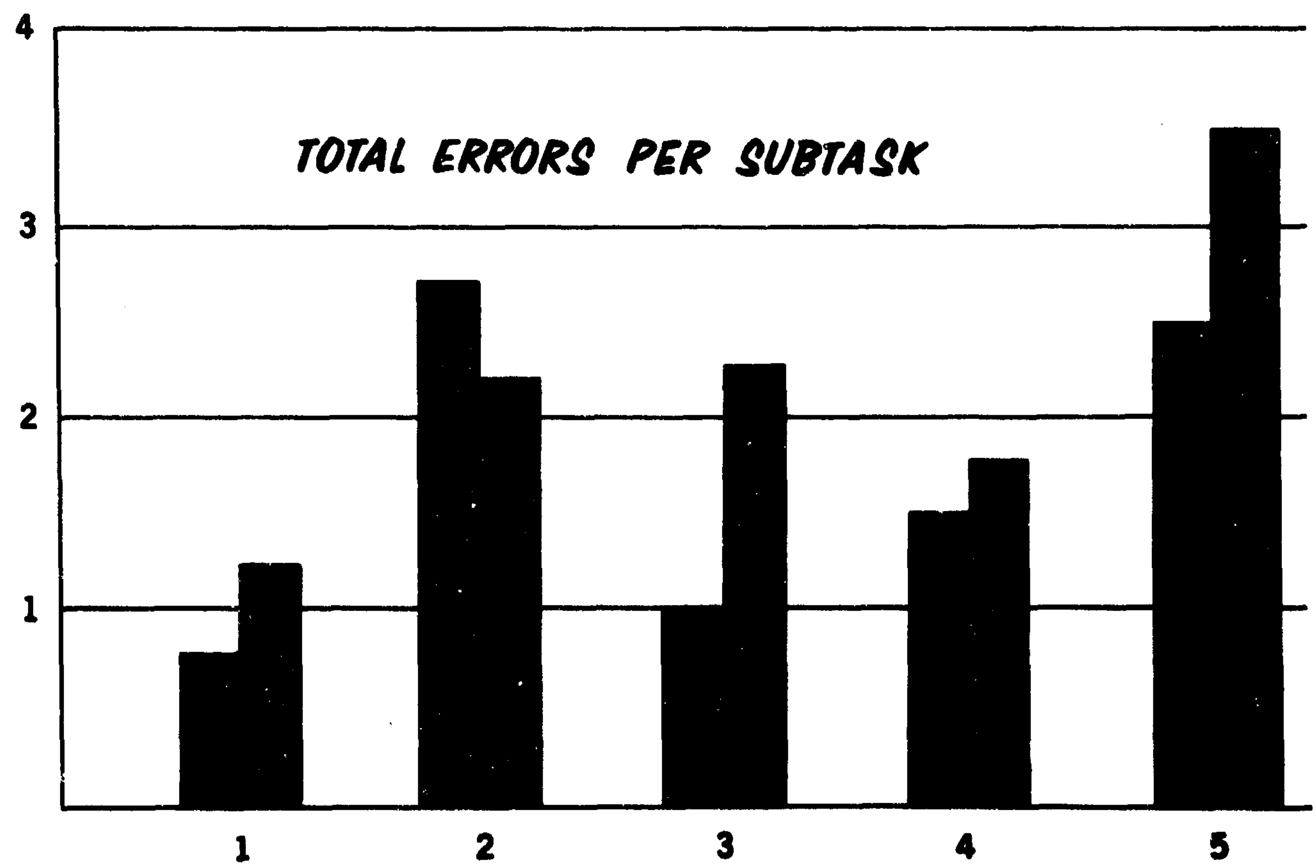


Figure 9. Bar graph of average number of total errors per subtask.

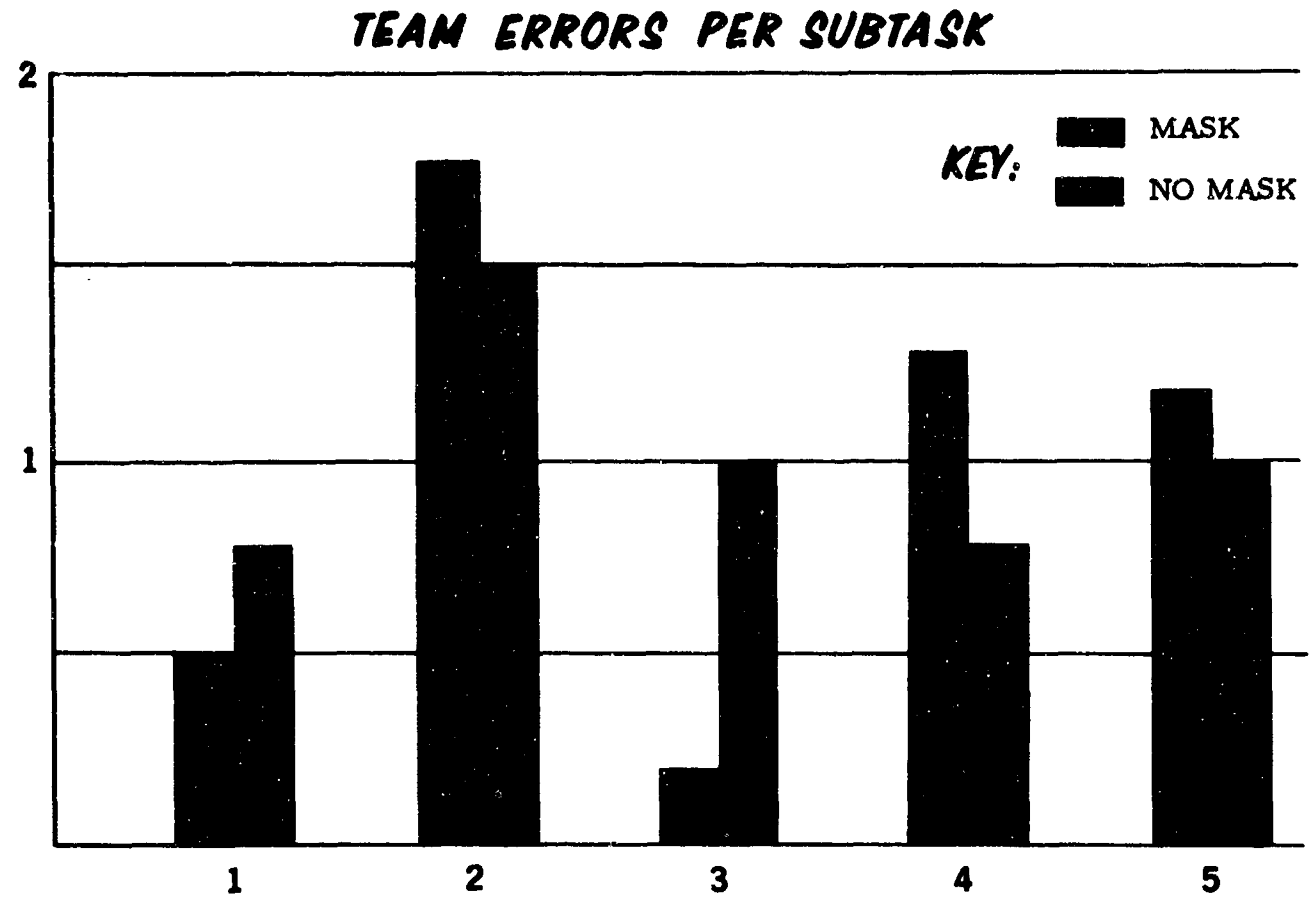


Figure 10. Bar graph of average number of team errors per subtask.

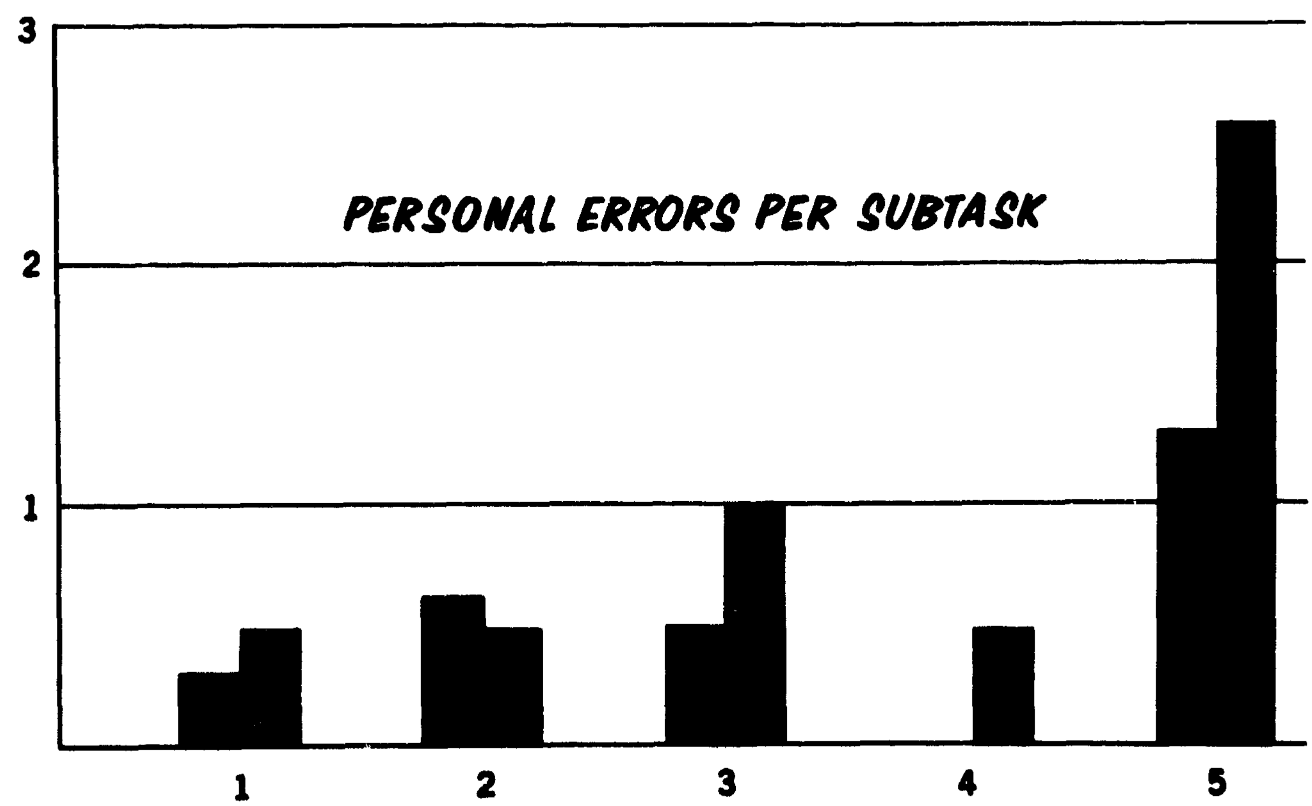


Figure 11. Bar graph of average number of personal errors per subtask.

difference between mask and no mask effects of the variable.

For example, total communications is the variable displayed in Figure 8. The general profile of the five bars representing the mask condition is consistently higher than the profile of the alternate five bars representing no mask condition. (This variable was, in fact, significantly higher with masks). The activity where this effect is strongest is in sub-task two. In this case, both greatest magnitude and greatest discrimination between conditions are found in the same sub-task.

This profile for communications, in which sub-task two is most significant, can further be compared with the profiles by sub-task of other variables. Personal errors, for instance, are displayed in Figure 11, and give a totally different profile--one that ascends with progressive sub-tasks, particularly in the masked condition. Here sub-task five yields the largest amount of data, but shares sensitivity to the difference in mask condition with sub-task three. A further analysis was made to determine which sub-task contributes most to the prediction of the total variance. Bar graphs for the other variables in this study are given in the appendix. The results, however, according to the category of variable and the strongest discrimination between the mask and no mask condition can be briefly outlined as follows:

(1) Oral intercommunication shows its greatest discriminatory effect in sub-tasks two and three. This involves the unscheduled verbals in sub-task two, and verbal repeats in sub-tasks two and three, with the mask condition exceeding the no mask condition.

(2) The visual intercommunication variable has its greatest magnitude in sub-task two, and next greatest in sub-task five. The greatest discrimination effect relates to the unscheduled visual and visual repeats in sub-task two with the mask. Discrimination between test conditions is also seen in sub-task five to a lesser degree.

(3) The total communications differ significantly in sub-task two. This is true for total magnitude and significance of mean difference; the without mask mean is 34.0 and the with mask mean is 50.3.

(4) Personal errors have predominant magnitude and difference in sub-task five. The greater effect is with the mask. Team errors are actually errors in training and coordination, and are shown by Figure 10 to be more frequent for the unmasked teams in the three sub-tasks that required knowledgeable individual activity--but more frequent for masked teams in the uncoupling and firing tasks which are always coordinated by the C/S. (This pattern may be typical of teams with minimal field practice; lack of initiative when masked actually precludes some errors).

Table H

Outline of Magnitude and Discriminability of Variables by Sub-Task

Variable	Magnitude and Sub-Task Without Mask		With Mask		Discriminability Between Conditions	
	Mag.	S. T.	Mag.	S. T.	d	S. T.
Sched. Orals	13	2	12	3	1.4	2
Unsched. Orals	6.3	2	9	2	2.7	2
Oral Repeats	0	-	.8	2	.8	2
Total Orals	19.2	2	20.5	2	1.3	2
Sched. Visuals	16.8	2	22.6	2	5.8	2
Unsched. Visuals	1.5	2	5.3	2	3.8	2
Visual Repeats	.5	2	1.8	2	1.3	2
Total Visuals	18.8	2	29.0	2	10.2	2
Total Comm.	34.0	2	50.3	2	16.3	2
Personal Errors	1.3	5	2.6	5	1.3	5
Team Errors	1.8	2	1.5	2	0.3	2
Critical Incidents	0.3	2,3	0.3	3,4	--	--
Total Errors	2.8	2	3.5	5	1.5	5
Quality	2.8	1	2.8	1	--	--
Time	3.8m	2	4.6m	2	0.8m	2

(5) Total errors take place mostly in sub-task five, but the whole task effect is not that pronounced; sub-task two shows more errors without the mask. All other sub-tasks have higher errors with the mask, with the greatest effect in sub-task five.

(6) Average quality rating produced no very strong average difference between conditions. Sub-task two was rated the lowest with and without masks, and sub-task one rated highest.

(7) The sub-tasks taking the longest time were two and five, with the greatest difference shown in sub-task two, where the mask condition caused the longer time. Sub-task two also predominates over sub-task five in magnitude as well as discriminatory difference.

Table H presents numerically the results of this sub-task analysis. The greatest magnitude of a variable is given in the first column along with the sub-task in which it occurred. The second column designates the greatest difference between the experimental conditions and the sub-task in which it occurred.

b. Regression Analysis

As an additional aid in evaluating the effects of particular sub-tasks, a series of multiple regressions were run on the SDS-920 computer to determine the best sub-tasks for predicting the total task data for a given variable. Table J outlines the weights found. This yields the sub-task which most effectively co-varies as the total score varies. In order to predict which sub-task could best estimate the total score, the sub-task with the highest β^2 weightings could be consulted. For example, the total visual and total oral scores are best predicted from sub-task two according to items 2 and 3 of Table J. (This also is the sub-task where the greatest magnitude and discriminability between mask and no mask condition exists according to Table H). For the total errors measure, sub-task two co-varies best with the total errors (although sub-task five has the greatest amount of errors and discriminability between conditions).

The results of the multiple regression analysis can be summarized briefly. Sub-task two demonstrates the significant effects of communication--both visual and oral. Personal errors predominate in sub-task five, team errors in sub-task two, and critical incidents in sub-task two, three, and four.

c. Density

The measure of density of activity was obtained for each man of the team for each of the five sub-tasks. This measure was calculated by dividing the man's activity time (time employed in actual job involvement)

Table J

Multiple Regression Analysis
Relating Sub-Tasks to Total Task Data

1. Total Errors

	1	2	3	4	5	Sub-task
Without Mask	0.21	0.58	0.01	.01	.002	
With Mask	.01	1.24	.08	.03	.01	

2. Total Visual

	1	2	3	4	5
Without Mask	.02	1.3	.00	.04	.18
With Mask	.03	1.7	.00	.00	.10

3. Total Oral

	1	2	3	4	5
Without Mask	.02	.20	.12	.09	.13
With Mask	.00	.92	.00	.00	.00

4. Total Quality

	1	2	3	4	5
Without Mask	.14	.43	.00	.75	.14
With Mask	.11	.08	.00	.47	.13

Magnitude of "weightings" indicate ability to predict totals scores.

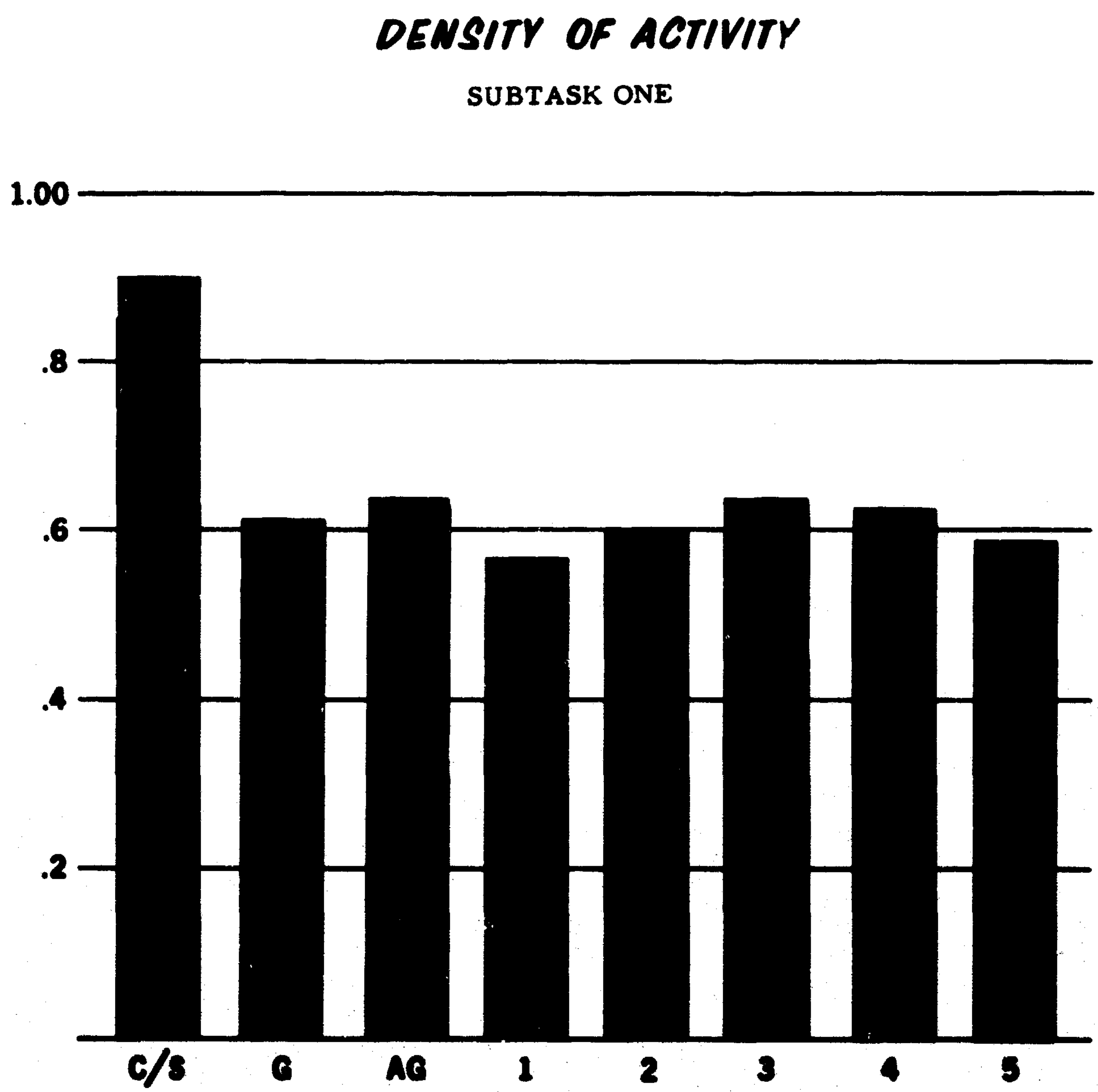


Figure 12. Bar graph of density pattern for team members. Red designates masked condition, and black the no mask condition.

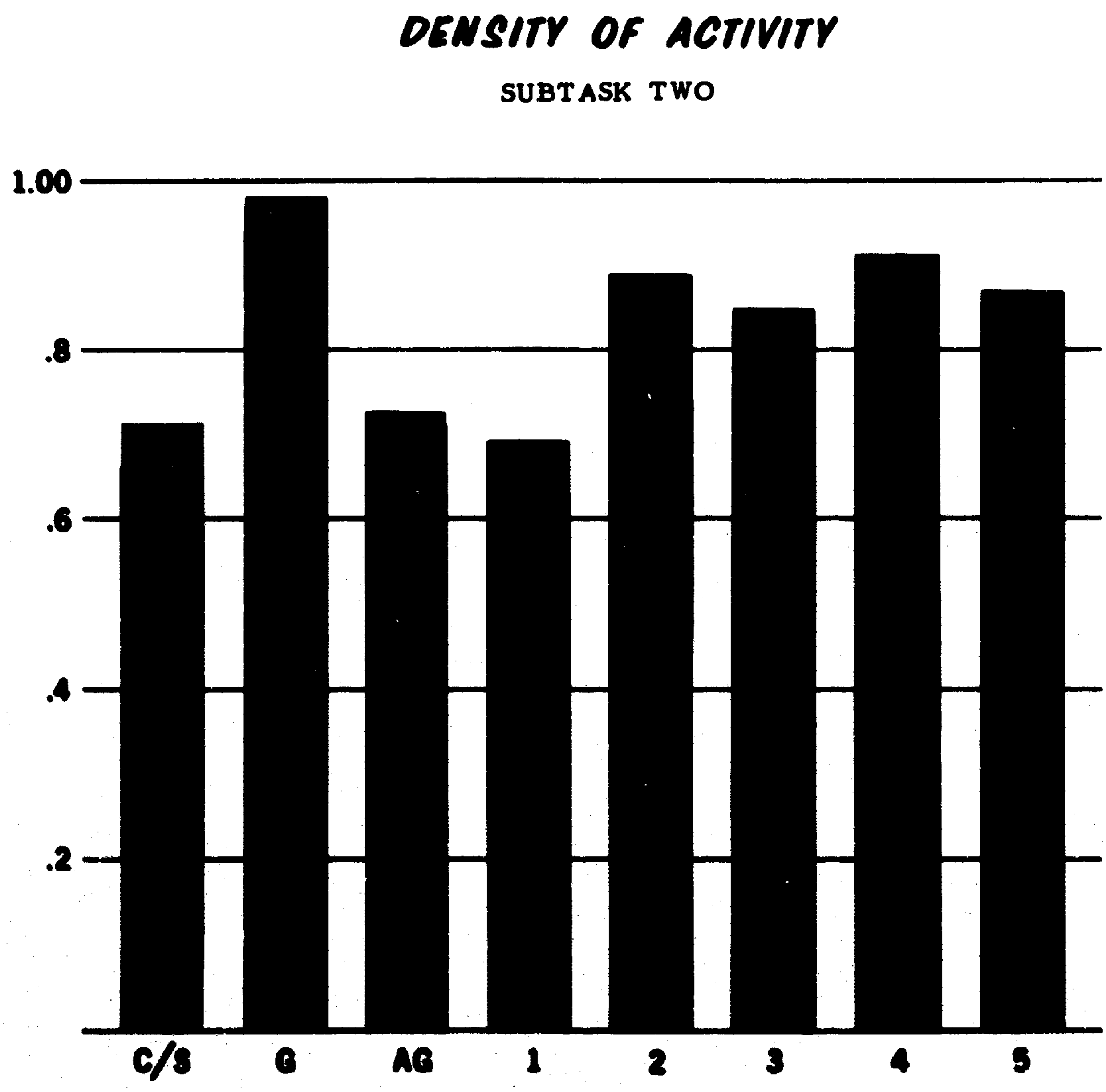


Figure 13. Bar graph of density pattern for team members. Red designates masked condition, and black the no mask condition.

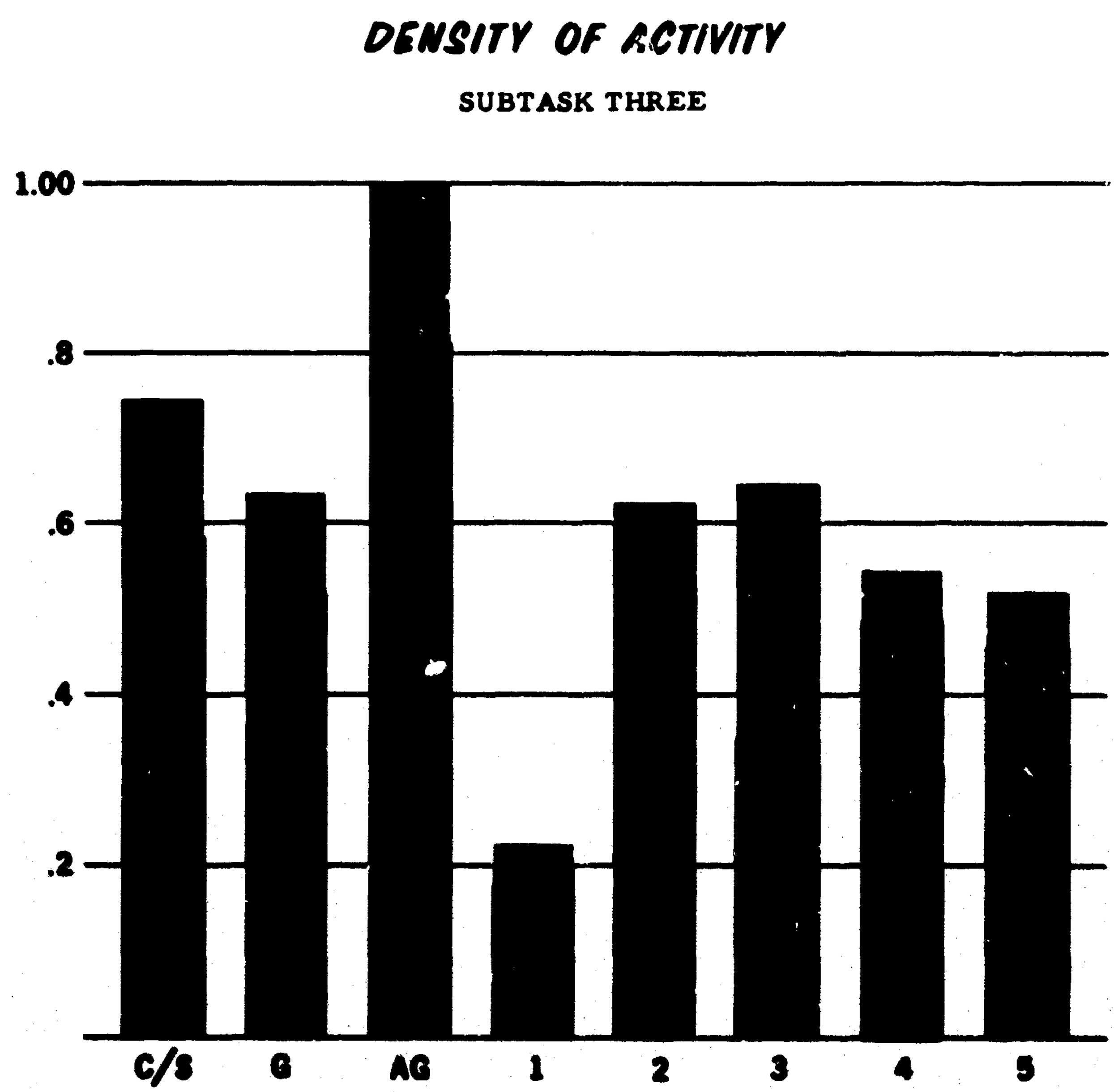


Figure 14. Bar graph of density pattern for team members. Red designates masked condition, and black the no mask condition.

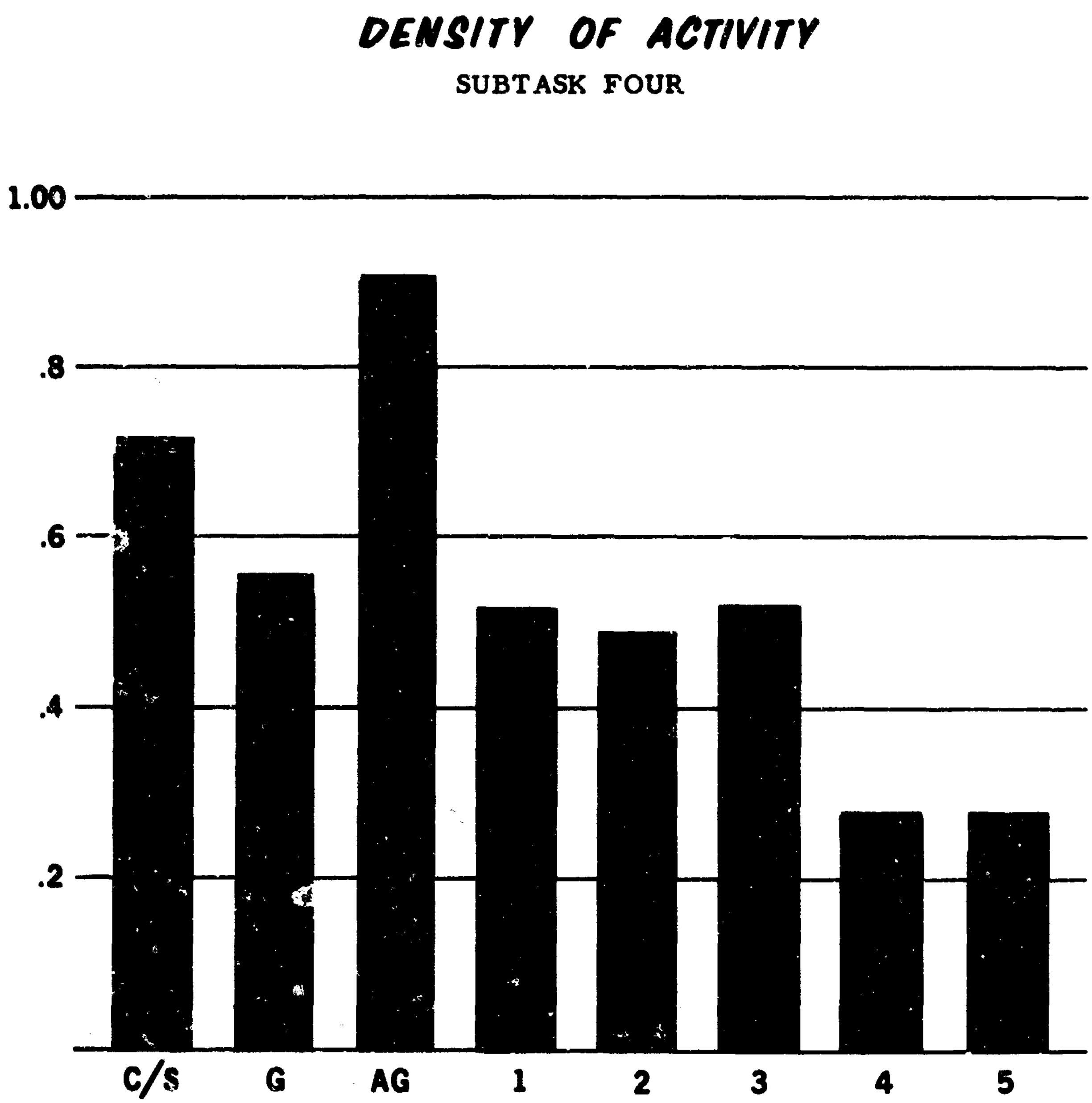


Figure 15. Bar graph of density pattern for team members. Red designates masked condition, and black the no mask condition.

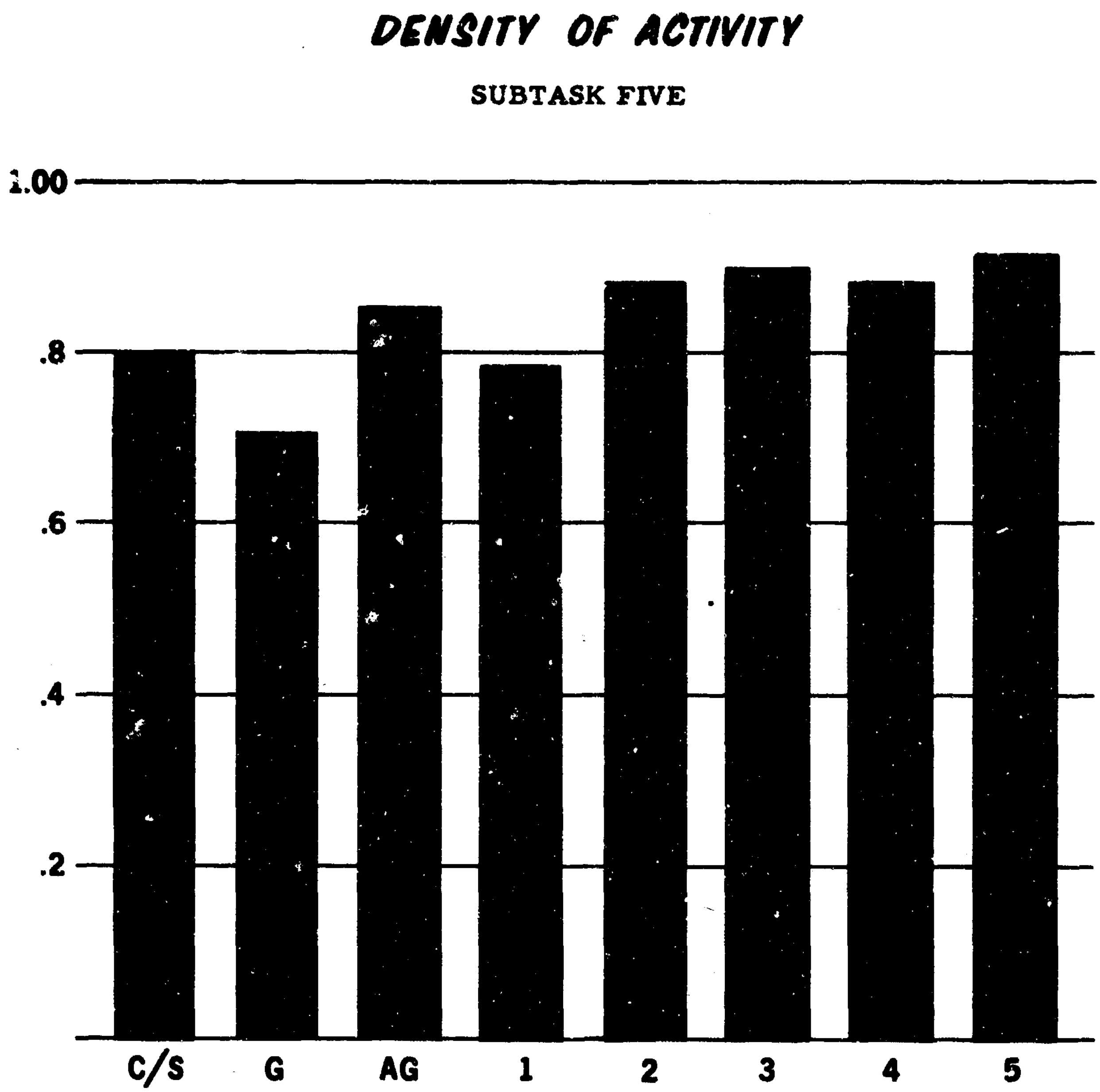


Figure 16. Bar graph of density pattern for team members. Red designates masked condition, and black the no mask condition.

by total time of the sub-task.

$$\text{Density} = \frac{\text{Active Man Time per Sub-task}}{\text{Total Time per Sub-task}} \qquad N = 6$$

The resulting ratio of density is for each of the eight team positions for each of the five sub-tasks averaged over all six participating teams. This was done separately for with and without mask conditions, and is shown in five density diagrams or "finger graphs," Figures 12 to 16.

In the discussion of the time variable on a sub-task basis it was noted that all sub-tasks took longer with the mask, the greatest difference being in sub-task two. If the density is the same with and without mask, it follows that the men are spending more actual time working with the mask on, since the total task time is greater with mask. In sub-task three, the density is greater with masks for seven of the eight team members; sub-task three also demonstrates the most difference between the patterns of density of activity under the two conditions. This sub-task is tactically important in that the explosive charge is set, the projectile fuzed, deflection and elevation adjusted, and the first round fired. It may be inferred that the men would be putting forth special effort to make certain this sub-task was done correctly, especially with the masks on. Sub-task four, where the following two rounds are fired, is the sub-task with the shortest time, since most adjustments have been made and the two rounds for effect are to be fired quickly. Here there is very little difference in density since the task is well patterned. Sub-task one, uncoupling the howitzer, shows the same effect of patterned team coordination. In sub-task two, the density is lower with mask for six of the eight men. The total time is much longer with masks in sub-task two, however, and this would tend to make the density less. The same case exists for sub-task five. The men individually are working approximately the same amount of time, but since the total team's time is longer to perform the task wearing masks, the density is less.

Presentation by sub-task density diagrams also reveals the relative amount of active man time among the team members. In sub-tasks two, three, and four, either the gunner or his assistant gunner are predominantly active, as would be expected by the nature of these tasks. In sub-tasks one and five, all team members perform at relatively the same level of density with the exception of the chief-of-section. His density is by far the highest in sub-task one, where initial coordination of the team is his responsibility, and quite low in sub-task five, in which the final "pick-up" work is done largely by the cannoneers.

d. <u>Heart-Rate</u>

The basic heart-rate data proved to be quite variable. Large

teams made monitoring of each individual not feasible in this study which concentrated on interactivities. However, by telemetering heart-rate from each man in a small team, a comparison could be produced of total work level against other runnings of similar tests, and also of relative work-loads to assure that no data was used from a subject "loafing" or overloaded in a given trial. In this study, each man's total task was divided into deciles, and the heart-rate for each decile was averaged over the five teams. This was done separately for the mask and no mask conditions. Graphs were then constructed on these ten averaged samples for the three men whose heart-rates were transmitted and recorded. These graphs are shown on the following three pages (Figures 17, 18, and 19). The often large alterations noticeable in the individual graphs taken from the recordings have been smoothed to give the general flow of the heart-rate during each sub-task (bracketed along the bottom). It is thus a general summary of the individual graphs.

Certain points are noteworthy in viewing the graphic results:

(1) The with mask condition is more variable, that is, the range of heart-rate is greater with masks.

(2) The number 3 cannoneer's graph is far more variable than those of the gunner or assistant gunner graph.

(3) The general outline of the graphs is what one would expect from the inspection of the task itself. That is, during the time immediately after jumping off the truck, the men are working hard, then there is a general lull, and at the "march order" command, work rate again increases as the team packs up to leave the firing site.

The effect of sequence appears to be reflected in the range of heart-rate. Those teams that performed first without a mask showed a narrower range of heart-rate with the masks during their second trial than did those teams which performed with the mask on the first trial.

Each graph can be noticed to follow basically the same pattern of rises and falls in heart-rate, demonstrating a reliability of the field data telemetry and recording. We may note that while the overall pattern of heart-rate does not differ throughout the task between the mask and no mask condition, the levels of heart-rate for the number 3 man diverge considerably in the two conditions during the first sub-task. No such apparent difference exists for the gunner or assistant gunner. *

* The number 3 man usually does more physical work in sub-task one than the gunner or assistant gunner. He runs up front to unlock the shield and position the axle lock, usually assists on the trials. Under this physical load, the mask effects may become more apparent.

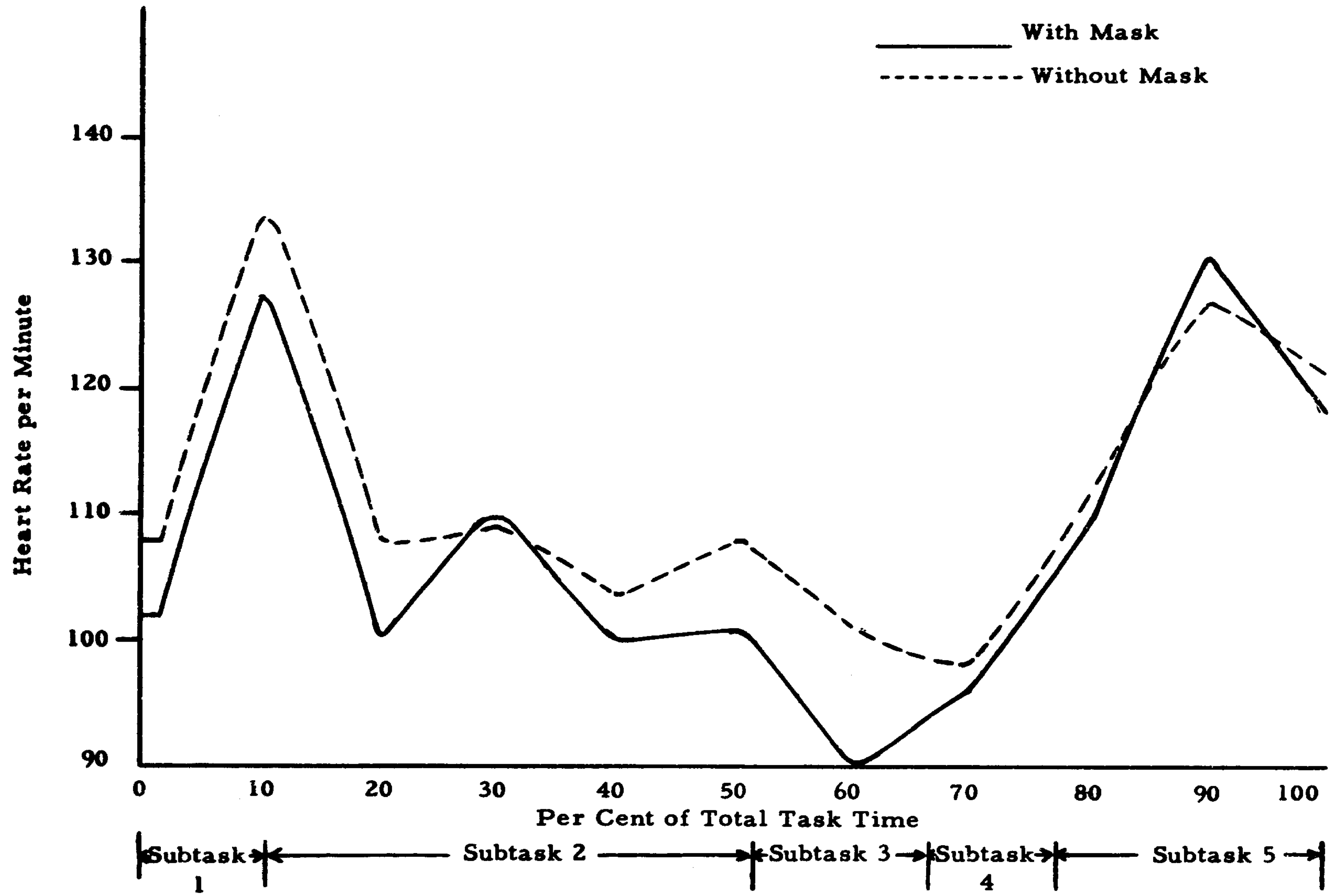


Figure 17. Average Heart Rate for Gunner.

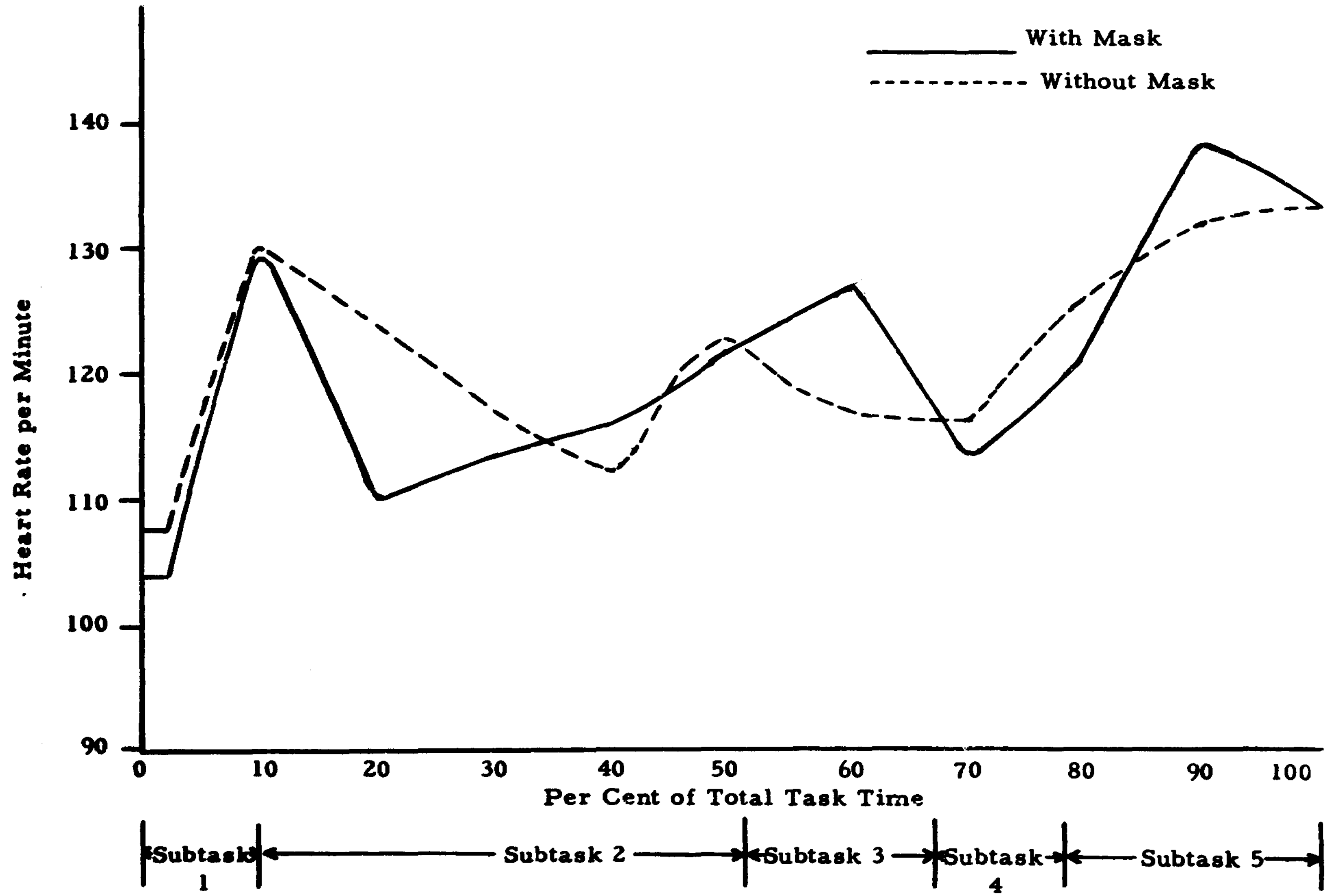


Figure 18. Average Heart Rate for Assistant Gunner.

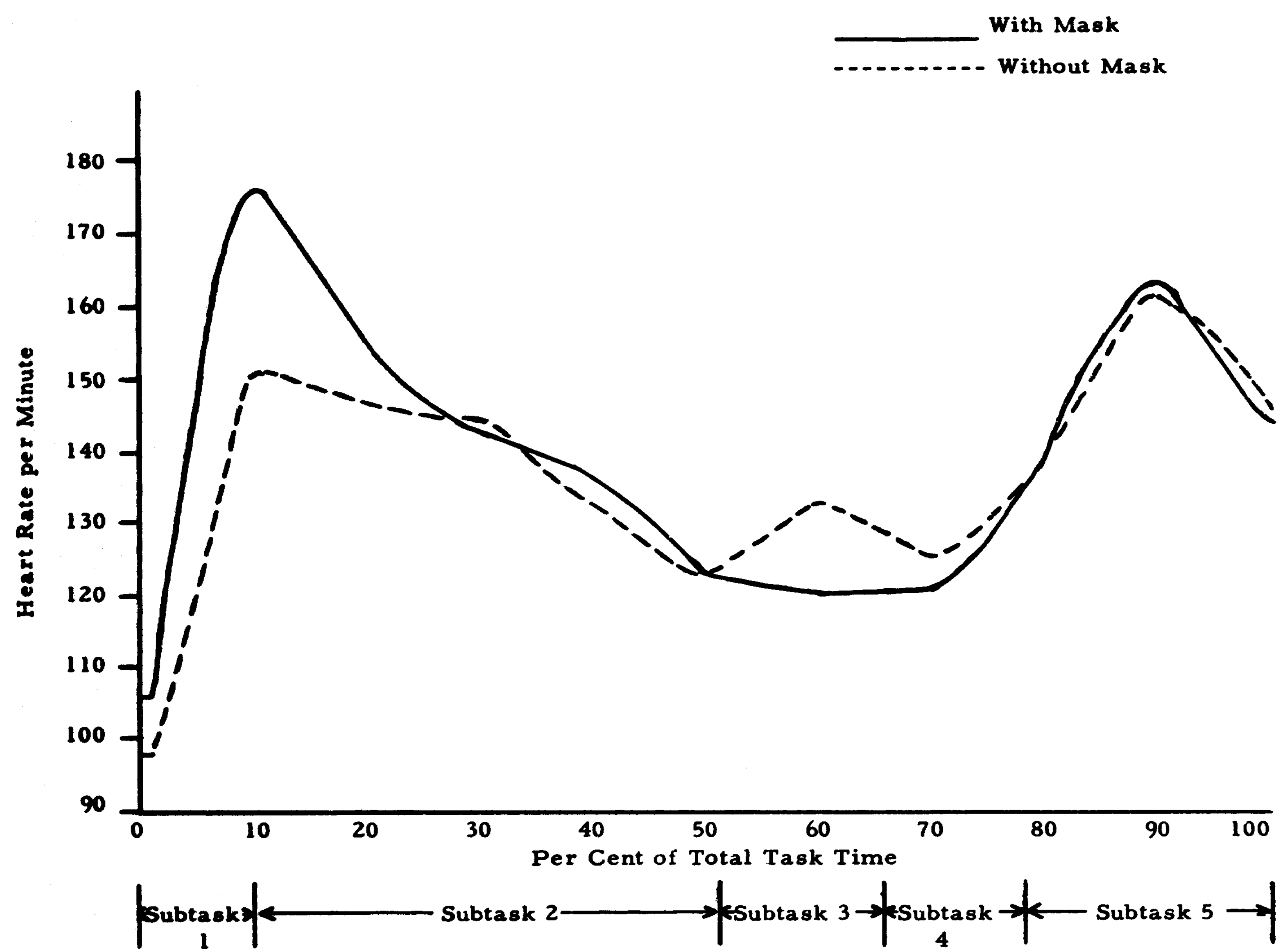


Figure 19. Average Heart Rate for #3 Cannoneer.

Since the tasks for each of the three men observed entail rather disparate assignments, no attempt was made to produce a summary graph of heart-rate for the three men. Observation of the three graphs sufficiently demonstrates the general pattern.

3. Reliability and Validity

Table K shows the obtained reliabilities of the variables. These are correlations between every other team for all teams participating, paired on a sub-task basis. To calculate this the first team was paired with the second, the third team paired with the fourth, and so on throughout all five sub-tasks. The two halves were then correlated for the variables shown in Table K. Reliability calculated in this manner is better called the "coefficient of data consistency." It demonstrates the extent to which one can depend upon the appearance of the variable in any sub-task. In general, the reliabilities appear to be higher for performance with no mask. Reliability also appears to be a function of the magnitude of the variable. The variables with substantial magnitude (appearing frequently in a sub-task) are generally those showing greater reliability.

No reliability could be obtained for quality, since the "good," "fair," or "poor" rating does not describe a certain level of variable with a continuous distribution. A correlation of this type would only describe the consistency of the rating between all the pairs of teams on the "good," "fair," or "poor" category across all sub-tasks equally, since the quality can never be greater than "good" (a 3 rating) in any sub-task. When the videotapes were reviewed, however, two separate observers rated quality. The last row of Table K reports the inter-observer reliability between them.

A validity of the measurement system employed in this study is drawn from a consideration of the sequence effects observed. (The tactical validity of the field performance is another question and beyond the scope of this study). The sequence effects were noted previously and can be seen in the line profile graphs. Table C outlines the negative differences obtained on the second performance of the howitzer task, regardless of mask or no mask condition. The teams show a gain in proficiency on the second run. An objective measure of this proficiency is the time variable. All teams took less time on the second time they performed the task. The negative values in Table C show that the other more subjective measures appear to be sensitive to this alteration in proficiency. Errors are less on the second run, and total deviations in general decrease on the second run. In general the measurement system is sensitive to changes in team variables in the same direction as the time measure.

Tetrachoric correlations were calculated for the variables according to sequence. The correlation between sequence and time on first trials and on second trials was .57. Although it was noted in a previous section that

sequence effects were not significantly different, they do co-vary with order of run. Other tetrachoric r's were calculated between time, as an assumed criterion of proficiency, and the other variables observed. The results were not significant, but the general trend is obvious in Table C. All the variables except one, scheduled visuals, change in the expected direction, pointing out that other than random effects appear to be operating. This internal correlation of measures seems to present a reasonable amount of metric validity. A more extensive independent investigation, however, would be required to obtain more conclusive validity data.

Table K

Reliability of Data: Coefficients of Consistency

	Without Mask	With Mask	Both Conditions
Sched. Orals	.71	.74	.72
Unsched. Orals	.86	.81	.81
Total Orals	.87	.70	.76
Sched. Visuals	.81	.71	.74
Unsched. Visuals	.29	.12	.20
Total Visuals	.76	.49	.58
Total Errors	.67	-.34 *	.20
Time	.86	.44	.63
Quality (inter-observer)	--	--	.84

* Negative coefficient because of fewer team (training) errors with mask.

C. Summary of Results

A small group tactical mission observed in the field formed the data source for this report. The resulting analysis pointed out variables relevant to small group interaction and performance which were able to discriminate differences between the performance of teams with and without the M-17 protective mask.

Team intercommunications were significantly higher for the mask condition. This effect existed most obviously in sub-task two, which demonstrated the greatest magnitude of the variable as well as the largest discrimination between the mask and no mask condition. Of the variables making up total communications (total oral and total visual signals), the visual signaling differed significantly with the mask. Visual signaling increased for all three

categories of the measure--scheduled, unscheduled, and repetitions. The total oral measure, on the other hand, remained about the same, evidently because in the mask condition scheduled orals decreased while unscheduled orals increased (organized verbal command gave way to ad lib verbal coordination). Table A summarizes these patterns, and also shows that with the mask the total scheduled communications increased only 7%, but the unscheduled communications increased 50%.

The measure of total errors increased, but not significantly, when the teams wore masks. The greatest contribution, however, to this difference in number of errors was from the measure of personal errors which concentrated its strongest effect and greatest difference in sub-task five. The pattern of distribution among sub-tasks proved quite different for the other categories of error. Team errors which showed only a small increase with the mask on, had the greatest magnitude in sub-task two, but strongest discriminative effect in sub-task three. Critical incidents were about equally distributed among sub-tasks two, three, and four; and had low magnitude. Teams uncertain of duties may actually show improved attentiveness to leader while masked.

The quality measure, as defined and employed, did not significantly discriminate between mask and no mask teams. In general, sub-task one was rated the highest and sub-task two the lowest in quality.

The time to complete the task increased with the mask on, not significantly, but consistently. Sub-task two took the largest amount of time, and also showed the greatest time difference between the conditions.

The heart-rate results on the gunner, assistant gunner, and number three man, showed no appreciably higher level (on the average) with the masks. The number three man demonstrated a higher rise in sub-task one when wearing the mask, as well as a greater variability of heart-rate than the gunner or assistant gunner.

The density of activity measure showed no outstanding discriminative difference between mask and no mask conditions. The density appeared to be generally larger, however, in sub-task two in both conditions. This measure did demonstrate the changing pattern of individual responsibilities within the team, inherent in each sub-task: In sub-task one (uncoupling), the chief-of-section is active a greater amount of the total task time; in sub-task two (preparations), it is the gunner; in sub-task three (firing), it is the chief-of-section and assistant gunner; in sub-task four (fire for effect), it is the assistant gunner; and in sub-task five (march order), it is the number two, three, four, and five men. Sub-tasks one and five showed the highest measure of density in general.

The effect of the hoods with the protective mask is very small compared to the difference between the mask and no mask. Only two teams were tested,

however, in this third condition, which appears to be only a slight modification of the mask condition.

The After-Task Questionnaires showed the men felt they could see, hear, and speak well enough to do their job. The only outstanding point of report is that the men seemed to feel more tired after doing the task with the mask on. The men showed a 15% increase in "worse" self-ratings with the mask on, although this is not significant. There was no noteworthy relationship between the men's subjective reports of performance and other external measures relating to performance.

A result, other than the mask and no mask differences, was the relationship between the order of a team's performing the task and the level of the resulting data. In all cases but two (quality and scheduled orals), there is an average decrement in the measure from run one to run two, with the greater decrement in the case of those teams who performed the task first with a mask and then without it. A series of profile line graphs were drawn to demonstrate this, and formulas derived to calculate the relative effects of the mask and of the order of performance. Because of the dominant influence of the sequence of performance, which involves both learning and mask adaptation factors, a careful separation of mask and sequence effects must be part of any standard evaluation technique; a recommended approach to this problem will be discussed in Section V.

D. Discussion of Results

The results obtained outline the directions certain variables will take when a small group tactical team wears protective masks while performing its mission. The summary of results given in the previous section presents the outcome of the data as it was analysed. Often this can be sufficient to evaluate a study. But it is advantageous in a developmental study to supplement the results with interpretive or expletive remarks.

This study's prime purpose was a design to detect pattern changes in team performance with masks, not validation of the tactical impact of such observed changes. Nonetheless, if masks decrease standard commands and increase ad lib verbal interchange among team members, then mission success is probably altered in some degree. We can deduce from the test analysis that if the intended pattern of communications is disrupted among the team it will not perform as effectively. This study was governed by the criteria of discriminative differences between the mask and no mask condition under the assumption that the variables chosen for observation would affect mission success. It would have been experimentally interesting to have intermediate criteria of mission success that could then be related to the variables observed on-site. Unfortunately, the more obvious intermediate criteria, such as accuracy of fire, involved other personnel besides the subject teams. The

determination of satisfactory "mission" criteria would be out of the designated area of this study. The extreme difficulty in assessing mission accuracy quickly and objectively is due to the many variables that enter into the determination of tactical accuracy of the Howitzer teams. For instance, the general reliability (standard error) of the Howitzer itself, the weather's effect on ballistics, accuracy of the battery observer, of computation of the fire control director, and abilities of the safety officer as he checked the readied Howitzer are relevant factors beyond the team's own performance. With so many components entering into firing accuracy, the measure becomes too cumbersome to be studied in assessing tactical accuracy of the team, except in large numbers of firings.

Concerning the value derived from heart-rate data, it may be inferred that the task in general was simply not strenuous enough to reveal any significant differences in the men's heart-rate in the two conditions. The number three cannoneer works harder physically in sub-task one than the other two men whose heart-rates were sampled. And in sub-task one this number three man's heart-rate is in fact higher with the mask than without the mask. That is, when the number three man is under sufficient load, a difference can be noticed in the heart-rate. The rest of the task is not really strenuous, especially with the light weight 105 Howitzer. Another point to notice is that the time was not kept constant and as a result, though the density diagrams show about equal ratios of active man time per sub-task, it must be realized that a trial time to perform while masked is longer. The men may be working harder, but it is spread out longer as well. If time were controlled, if the masked team were forced to complete the task as quickly as when not masked, the difference might show up in a higher heart-rate.

The quality measure did not demonstrate any apparent discrimination between the conditions. One might expect that if intercommunications increased significantly with the mask, this would affect mission success. Why is it that the expert observer of quality did not discriminate between team conditions in his ratings? In this case, the nature of the metric used may offer an answer. The observer only rated on a three-part scale, "good," "fair," or "poor." This possibly did not allow enough range to produce a difference. The observer did rate sub-task one highest, where the average error rate is less, and he did rate the second run slightly higher than the first run on the average. This may lend some validity, even to the three part measure. In future studies, a more expansive scale would be recommended; either a continuum along which the observer could rate subjectively, or a ten point scale with observer standards to be met for each progressive rating of the performing team.

In the study it was noticed that the unscheduled communications increased with masks. It may be expected that the scheduled measures should remain approximately constant, since there is only a limited number of them defined

for the activity and all are considered necessary. For the oral commands this was true. Certain commands must be given orally, even if it takes more effort with the mask on to do so. Even so, a slight decrease in scheduled orals is noticed in the data on the masks. Scheduled visuals are free to increase since certain visual signals are not limited by a standard procedure. In setting out the aiming-posts, the gunner can signal visually as much as he needs, or feels he needs, to align the posts properly. And if speaking with the mask is more difficult, it would be expected that the men might naturally fall back to visually signaling to make their communication more clear. The increase in scheduled visuals may be interpreted as extra measures to make sure a communication is effective while wearing the mask. In the case of oral communication, the unscheduled oral signals increase with the mask, which may be a further indication of additional attempts to make clear a communication to another man, either because of difficulty in speaking with the mask or because of a degree of uncertainty and isolation that it produces within each subject.

A few aspects of counterbalance in experimental design were brought out in this study. The purpose of counterbalancing the presentation of the experimental variable is to cancel out the effects of sequence on the experimental condition. For small sequence effects, and a fairly large sample, this usually works out in good order. But where the effects of sequence appear to be large relative to the experimental effects, this arrangement can produce a problem. When learning effects are equal, the counterbalance design still yields the same mean difference between the experimental condition and the controls, but the standard deviation may be grossly affected: The greater the decrement in data from the first to the second run, regardless of initial mask condition, the larger the range of the variable. One might even expect a bimodal distribution in an extreme case--one area of distribution about run one and the other about run two. This effect makes it very difficult to achieve statistically significant results. This order effect is very strong on many variables of the study, and a few are not distributed homogeneously. When the sequence effects are so strong, the mean difference will not change, unless there is the additional problem of unequal sequence effects. In either case, it is suggested that when sequence effects predominate over experimental effects, the two be isolated by a presentational technique such as Figure 4 and analyzed separately to allow a further understanding of how the variable in question is behaving in the study, such as has been done here by the derived formulas concerning mask versus sequence.

Although they have generally been called "sequence effects" thus far, since the sequence of conditions alters the resulting data, it may be inferred that there is actually learning underlying the effect. The variable of time, for instance, shows a marked decrease on the second run regardless of the experimental condition. If time is one measure of proficiency, the teams have learned to be more proficient. Indeed, it would be expected that the

teams should learn in this situation. The teams repeat the performance in the same spot, the Howitzer is fired in approximately the same direction, and the same men repeat the same jobs. The man setting out the aiming post even runs to the same spot on the second trial. A learning of the "task in this situation" is then most probable. This would involve several variables of the task performance behavior, of facility in communication among all team members, and of adaptation to the test site and situation. The sequence effects are most probably such learning effects.

These sequence, or learning, effects were first viewed as a variable hiding the experimental variable. This did happen, but further analysis allowed compensation and a relatively objective view of the experimental conditions of mask versus no mask performance. The learning effect even produces some benefit by permitting an internal validation of the measurement system used. The internal criterion used was gain in performance, and there are objective measures of this. Time is an objective measure which should, for reasons previously explained, belie a team's gain in proficiency. If the performing teams, on the average, cut their time by 2.3 minutes the second time they performed the task, they would certainly seem to have become more proficient. Aiming-post time is even more illustrative of this. Along with this gain in proficiency, it can be seen, in Table C, that all other measures demonstrate the same trend because of the negative quantities obtained under sequence effects. The only positive results are quality and scheduled orals. The quality should be expected to increase and so should scheduled orals, since many are designed to aid proficiency rather than being absolutely necessary. The correlations between the proposed internal criterion of time and other measures of a more subjective nature were not calculated, but inspection of Table C offers support for the validity of the measurement system used. (It may be noted also that the observers viewing the videotapes were unaware of exactly which team or what actual field sequence they were observing at any given time). A more detailed study is suggested to better assess validity of the measures obtained.

On the After-Task Questionnaire, the men reported that they felt tired after wearing the mask. This seemed to be their strongest response to the mask and is consistent with findings of previous mask studies. The Oral Interactivity Observer, who handled the After-Task Questionnaire, observed a general appearance of overheatedness around the face at the completion of a run with the mask, even with field temperatures near freezing. The feeling of tiredness may be cued in part by the perspiration in the face as well as by the increase respiratory pressure. The Howitzer teams that participated in this study were all in training and their reference for normal performance would be expected to be affected by their level of experience. That is, their idea of normal performance and an observer's point of reference may be on an entirely different base. As a result there is no relationship between the men's subjective assessment of their performance

and the observer ratings.

Team and personal errors are higher when the men wear the mask. On the After-Task Questionnaire, however, 85% of the men reported they could see well enough to do their jobs. It seems personal errors, especially would be a result of not being able to see well. Errors such as tripping over the trails, tumbling into a hole, or bumping into another man, would certainly imply a man can't see well enough. Brief matching of individuals responding against those committing the errors reveal no clustering of errors around the remaining 15% of the men who did not report satisfactory vision. The point made for discussion is that errors in general would seem to be caused by a restriction of vision due to the mask.

The reliability correlations shown in the results are better viewed as for dependability of appearance of the team evaluation variables in any sub-task throughout the whole task. In general, the without mask condition demonstrates a higher dependability that a variable will appear from team to team as runs progress. The with mask condition apparently causes a disruption in the appearance of a variable--that is, its count per sub-task becomes erratic. We may infer, based on this, that the pattern of behavior of the men as a team has become more erratic or disrupted with the mask on. The reliability, obtained by the split-group method, can be expected to be higher with greater magnitude of the variable being observed. The variance of a low level measure is more greatly affected by a single change in unit count, than the variance of a measure greater magnitude. This fact is noticeable for this study in a comparison of the bar graphs and the reliability shown in Table K. The inter-observer reliability was not available for all observers. Each observer was generally doing his own particular job in the field and also when observing the videotapes. On quality, however, two observers were scoring much of the time, and this is shown as an inter-observer reliability.

The multiple regressions were computed in order to allow a view of which sub-task co-varies most with the total of all sub-tasks or with a particular variable. A sub-task may show the greatest magnitude of a variable or the greatest difference between the mask and no mask condition, and yet not co-vary at all with the totals data. This information was useful to this study for developing recommendations for standard test evaluation and analysis.

Gathering the data through the prime medium of videorecording allowed a comparison between the relative yields from the field-site and the video-tape viewing. Many variables were gotten only from the videotape, such as personal errors, team errors, and the times. But other variables, such as oral communications, were obtained in the field via observer checklist and also from the videotapes using slightly modified checklists.

The difference in magnitude of the yield proved to be 34.3% greater from the videotapes than from the field. It must be realized, however, that the data gathered from the field and the videotapes is not totally isometric. For the videotape data the variables were more definitely defined, or slightly redefined, in view of what was learned from the field. The greater amount of data obtained from the videotapes can be attributed to the less distracting atmosphere while reviewing the tapes, as well as the ability to slow, stop, or replay the tapes when an observer is in doubt about data relevant to an activity. In this way videorecording provides a unique advantage over the actual field trials in data handling and the usable yield is greater.

SECTION V

DISCUSSION OF EVALUATIVE TECHNIQUES

All measurement techniques investigated in this study yielded useful data when applied in the field, but none was an unqualified success. Of the seven techniques field-tested, three are direct recording: videotaping, filming, and telemetry of heart rate (all continuous during team test activity). Two techniques are indirect recording of observer-detected data: the Data Sheets and the audio taped notes. The remaining two, the post-test questionnaires and the unscheduled Super 8 and 35 mm photography, are attempts at representative sampling, subjective and objective respectively.

Ranking of these seven techniques by their actual use throughout tentative design, field observation and evaluative analysis during this particular study would result in this ordering: *

- Videotapes
- Observers' Data Sheets
- 16 mm Films
- Audio Taped Notes
- Telemetry Charts
- Post-Test Questionnaire
- Super 8 and 35 mm Photos

Although it cannot be demonstrated that these techniques would prove of the same relative values in all field evaluative tests of human performance, their ranking in this general order of usefulness for any team performance evaluation of similar field design is strongly indicated by the

---

* Note that those techniques resulting in a continuous monitoring of the activity are ranked highest and the two sampling techniques are ranked lowest. Where data collection in depth is desired, such as in an analysis of subtle and complex team interactivities, recorded data in its fullest form will generally be preferred. Only if underlying differences were to be purposely ignored, such as in the "either/or" determination of mass voting tendencies, would gross sampling techniques be preferable to fuller data recording.

experience encountered with each. The following paragraphs analyze the inherent and incidental advantages of each evaluative technique noted during the Ft. Bragg observations, as well as the disadvantages encountered. Since no irrelevant malfunctions of equipment nor radically abnormal events occurred to disrupt this observation exercise, it must be assumed that the characteristics here described would probably be typical of small team performance evaluation at a single field site.

A. Videotaping

For the analysis in depth of patterns of team interactivity required in this study, videotaping proved to be the most valuable and trouble-free technique. Its use for field data collection presents several distinct advantages:

- Synchronized display of running time with picture.
- Continuous recording of team performance for up to four hours.
- Multiple-perspective view of activity.
- Simultaneous recording of visual and verbal aspects of performance.
- Additional verbal recording capability for running commentary.
- Non-interference recording from relatively remote distance.
- Adequate recording under marginal field conditions of light and weather.
- Erasability of irrelevant or disqualified data for reuse of tape.
- Immediate on-site monitoring of visual record for improving techniques.

Several disadvantages are also apparent, although all proved minor in this study:

- Initial capital cost or leasing cost of portable video-recording system.
- Availability of experienced TV technician and/or cameraman.

- Mobile van or panel truck to house supporting electronics.

- Requirement for a regulated source of electric power in the field.

- Need for coaxial cable connection between cameras and truck.

- Later availability of a video playback system for viewing tapes.

Following is an analysis of these major characteristics of videotaping.

Advantages

1. Synchronized Display of Running Time

The electronic insertion of an image of a minutes-and-seconds meter into the upper left corner of the picture, both on "wide" view and "detailed" camera recording tapes, was the greatest single aid to data collection and analysis. It allowed tagging of specific events for repeated viewing, for viewing from the detailed perspective, and for comparison against film records, field notes, or other data sources to be integrated without duplications. The visual presence of the time readout greatly facilitated precise timing of all team and individual tasks, sub-tasks, and activities without need for use of stopwatches at the test site. Such insertion of time at the field site would have been tedious or impossible on filmed recording.

2. Continuous Recording of Team Performance

The task monitored in this study had a duration range of approximately eight minutes to twenty-four minutes; or a ratio of one to three between shortest and longest run. Filming continuously for twenty-four minutes is not feasible with most photographic cameras for field use--particularly at sound speed, as desired in this study for later addition of sound-track commentary. In addition, the accumulation of sequential test runs, more than ten hours of recording in this instance, is much more convenient on videotape than on separate reels of film which must be spliced into several large reels after chemical processing. In fact, by erasure of "dry runs" from the videotape while at the site, the entire official sequence of observations for this study was contained on only two four-hour reels, one for closeup and one for "wide" observation of the entire team.

3. Multiple-Perspective Observation

In the uncontrolled field environment, a significant activity does not always take place conveniently in one location, within the clear and unobstructed line of sight of the responsible observer. Use of visual recording from two or more angles usually assures an observable view of an action that may be blocked or confusing when seen from the primary angle. After the scheduled three observation "passes" or viewings through our wide-shot videotape, several additional errors and team interactivities were detected, in later observations, from additional viewings of the closeup videotape scenes and the filmed recordings taken from the rear and left of the howitzer respectively. With videorecording, there are unique capabilities for improved coordination of the observations in the field at a central display where the activity can be shown from several angles at once, and later, for improved detection and interpretation of personnel activities by simultaneous monitoring from both close-in and overall viewpoints. These different views can be superimposed on one screen, electronically "split" onto two sections of the screen, shown in synchronization or overlapping sequence, or presented on separate screens. For example, a specific team activity can be shown from the "wide" perspective, orienting the observer to details to which he should give attention; then the same sequence can be immediately repeated from the closeup videorecording to note these details meticulously.

4. Simultaneous Verbal and Visual Recording

The synchronous availability of visual and aural cues to an observed activity was surprisingly helpful in the detection and identification of performance data. The on-site observer takes this for granted, but an observer of silent film or of recorded audio alone becomes well aware of the limitation of each by itself. Where precise speaker and hearer identifications are desired, as for intercommunications evaluation in this study, synchronized audio and video are essential to the observer not at the field site, who then has the added advantage of replay in case he cannot keep count of the pattern of verbal interchange at real-time pace, a distinct difficulty with the unscheduled verbals of the inexperienced teams observed. Conversely, when one of the simultaneous data distracts the observer from the other, as with the boom of the howitzer being fired, a simple turn of volume or video knob momentarily removes the distraction and allows reliable monitoring without the startle reflex. In at least two instances, TV observers who had also been on-site and noted no discrepancies in the brief seconds following the near-deafening discharge of a round, later detected deviations in the performance of the individuals preparing to reload the howitzer after that round.

5. Synchronized Commentary Capability

Although personal audio recorders were used by each field observer, they lack the tape capacity and battery power for continuous recording throughout the duration of the test runs. An additional audio channel available on the videotape, however, allowed the Team Events Observer to make a running commentary on the activity from a hand-held field microphone and provided perfect synchronization of remarks with the recorded action throughout the observations. An improvement on the capability was discovered almost accidentally during evaluative viewing of the videotape. By mistake, the commentary for some runs was recorded only on the closeup videotape reel. To hear it while viewing the wide-shot reel of tape, this commentary was transcribed to a reel of audio tape to be played along with the videotape. In attempting to synchronize the replay of these two tapes, the audio was set slightly ahead of the video, so that the observer's commentary on an event was heard a few seconds before the event actually appeared in the picture. This audio "leading" of the observed activity served to alert the observer of the video to coming details and to facilitate considerably their detection and evaluation. If the field remarks of additional observers are desired synchronized with the visual record, additional microphones can be "mixed" into the commentary track, or multiple audio tracks created if separate monitoring is a requirement.

6. Non-Interference Data Collection

Since the primary justification for field observation as contrasted to laboratory observation is the evaluation of the "real-life" factors in a performance, it is critical that the field presence of data collectors bring minimal distortion to the normal environment in which the observed personnel are to perform. For this reason, observers at the Ft. Bragg field exercises dressed inconspicuously and spoke quietly during activity, and the Team Events Observer, whose duties required his placement only a few feet away from the participants, actually wore olive drab fatigues and an army jacket. Recording equipment is generally believed to create considerable disruption in normal behavior of those "on camera," and, in fact, the total of six cameras, four observers, three telemetry pick-ups, and two microphones did apparently cause such self-consciousness in the initial team that the first "dry" run was performed in virtual silence, and in slow motion, so that the running time was more than twice what later proved to be typical. However, this initial "stage fright" phenomenon was largely dissipated, without any identifiable distortion of data on later "official" test runs. More significantly, the TV cameras seemed to become the least obtrusive of the data collection devices as the tests progressed. The unmanned wide-shot camera operated silently and continuously from a fixed pedestal far outside the radius of activity. In field recording by T alone, there would be no need for over-the-shoulder observers; a camera with a telephoto zoom lens could "move in" on the detail of a particular critical event. On-site coordination could

be done from out of sight, inside a video display truck at a remote location, and movement of participants across several dozen yards could be followed in detail by simply "panning" the telephoto camera and a remote directional parabolic microphone, rather than running alongside the participants. The psychological "presence" of "scorers" could, in effect, be largely dismissed from the minds of the participating team through the routine use of videorecording in the field.

7. Adequate Recording Under Marginal Conditions

The weather during the Ft. Bragg observations was windy, cloudy and cold. The below-freezing temperatures interfered with the handwritten note-taking of observers, with the placement of heart-rate electrodes on the subjects, and with the battery power supplies of the portable audio recorders. The gustiness made Data Sheet clipboards unmanageable and caused constant changes in the cloudy-bright lighting settings of the color-film cameras. The television system accepted lighting fluctuations much more successfully, was sufficiently warmed by its internal generation of heat to adapt to far lower field temperatures than encountered, and also provided shelter within its truck for the recording technician and monitors of the displayed activity. For special-purpose field observations, TV pick-up systems exist that can brightly display activity taking place in virtual darkness and amplify sounds and whispered remarks for easy audibility.

8. Erasability of Videotape

In addition to the two dry runs, several test runs were discontinued or later disqualified by the field Test Umpire because they fell far outside the defined limits of field performance to be observed. (One team, believing itself in another dry run, neglected to load and fire actual ammunition). The recordings of these runs were thus of no further value. The exposed color film and used telemetry graphs could not, of course, be reused. But the videotape holding these runs could simply be rerun for subsequent magnetic recording. Similarly, the entire eight hours of videotape can be erased during reuse at a later date, after all the data has been obtained from it. Although the price of videotape is now only a very few dollars an hour, this erasure capability, together with the absence of direct processing costs, allows the long-term cost of videorecordings to be reduced impressively in evaluative applications, where no permanent record is ever desired. This low cost encourages a freedom in recording large amounts of field data that cannot fail to be beneficial for in-depth observation and evaluation of the activity.

9. Immediate On-Site Playback of Recorded Data

Preliminary design and preparations for field observation can never completely anticipate the actual conditions of the observations. More

significantly, even subjective appraisal during observation can be erroneous to a serious degree. The availability of instant replay of the videorecording of an observed activity places the evaluator one step removed from the field situation for a more objective look while he can still do something to modify or improve the data collection technique or the scheduling of the observation structure. At Ft. Bragg exercises, for example, replay of the early runs resulted in revised instructions to a film cameraman to cover activity of the No. 3 man at the front of the gun, to the Team Events Observer to speak lower and give fuller running description, to the closeup TV cameraman to use a "tighter" lens on certain details, etc. Field replay was also a simple direct check that the fixed camera scene was including a view of all team members, that there were no technical flaws in the audio or video recording, that identification numbers on the vests would be legible to later observers of the tape, and that the angles from which recording was being done had good lighting and sightliness. Replay of an initial run*, moreover, is a convenient means of better orienting the field observers with repeat practice in the observation without involving tactical participants and equipment in further dry run activity only to train the observers.

Disadvantages

1. Cost of Videorecording Systems

Because of the already extensive integration of videorecorded laboratory experimentation into on-going projects at Edgewood Arsenal Research Labs, the budgeting of television facilities and equipment for this study's field evaluations was routine. Many organizations may not find the TV technique so readily available. However, this should be a serious barrier for only a short time longer. Initial production model videorecorders in the late 1950's sold for $45,000. Later models were produced for $15,000, in the early 1960's and recently, high-quality videorecording systems, including cameras, have become available for a little more than $5,000. Two lower-

---

* The incidental use of field replay of a team's performance, at the request of its supervisors, as valuable "mirror instruction" in the correct and incorrect procedures just committed, is an added and apparently quite effective training feature. If done, however, it should be systematically done for all teams after all runs, or in a pattern carefully planned to counterbalance all effects of increased knowledge of performance and competitive motivation with which it may endow a team for its second or third attempt at the subject task. At Ft. Bragg, informal viewing of replays by each team may have possibly contributed to the pronounced reduction in task time and errors by certain teams, regardless of mask condition.

quality models, "home videorecorders," are now distributed at little more than $1,000; and several companies are in competition to market a system for only a few hundred dollars within two or three years. Leasing of the recording system is also feasible if it is known to be a once-only use. As noted previously, because of absence of direct processing costs and the erasability of the tape, long-term use of videorecording is probably already less expensive than 16 mm filming for "evaluation-and-discard" purposes.

2. Availability of Experienced Video Technicians

For the Ft. Bragg observations, a company specializing in television recording was subcontracted for the week in the field. Television recording technician and cameraman were provided. However, except for the maintenance of the equipment, which is minimal on new units, a trained TV engineer is unnecessary once basic skills have been acquired in the manipulation of simple camera, recorder and playback controls. Again, if the once-only hiring of a technician and/or cameraman is indicated, they are more easily found each year in the rapidly growing number of closed-circuit industrial, commercial and educational TV companies.

3. Mobile Television Van

The large and expensive television vans used by commercial TV stations and networks for "remote" telecasts are no longer needed with the much simpler portable recording systems adequate for evaluative use in the field. The entire equipment complement, including cameras and cables, can be transported in a small panel truck or large station wagon. A check with Firing Range Control at Ft. Bragg located an appropriate site for the exercise that was accessible by reasonably smooth graded roads. No technical difficulties were encountered with the television equipment, even though it had to be driven to the site, unloaded, set up, used and "broken down" each of the five days of the field trip.

4. Requirement for Regulated Electric Power

Most portable television recording equipment operates on standard 110 volt a.c. current and uses low amperage. However, it must have a regulated voltage source to prevent imbalances in the electronic circuits. The selected site at Ft. Bragg had a tie-in box to the Post's commercial power supply. At field sites where commercial power is not available within a few hundred feet, a small, trailer-mounted gasoline generator can be used.

5. Coaxial Cable Link to Cameras

The necessary cables from cameras back to the television truck can be cumbersome in the field, and do limit the mobility of the cameras.

During one run of this study's exercise, a member of the tactical team dug his pit for discarding unused powder charges within inches of the coaxial cable. Since the cameras cannot be successfully moved over rough field terrain in any event, the loss of mobility caused by the cables is not serious if original camera positions are well laid out in relation to the anticipated activity. Links of fifty or even one hundred feet are frequently used, so that distance from the TV truck is no great problem, unless cameras are desired more than two hundred feet from one another to achieve radically different perspectives on an activity. The cables are reasonably rugged and weather-proof, but should either be dug in or laid out of the path of personnel and vehicle traffic to prevent damage.

6. Availability of Videotape Playback Units

Unfortunately, use of videotape is not yet widespread enough that video playback units are commercially available separate from the recording system. For the after-task viewing of the Ft. Bragg videorecordings, Dunlap and Associates rented this equipment from the subcontractor, I. T. V., Inc., by the week. Unless the system has been purchased for permanent test and evaluation use, a similar solution will usually be the most feasible for the foreseeable future. Since rental, although not prohibitively expensive, is not low-cost, it is advisable to draw up in advance a schedule of viewing "passes," number of hours or days required for each, and assignments of data collection and reduction duties to each of several simultaneous observers of each of these "passes" through the videotapes. The viewing schedule for this study involved three passes through the tapes, each viewing by four observers equipped with stopwatches, data format, and a briefing in the test design and objectives. This advance preparation, in addition to making after-task data collection more consistent and efficient, undoubtedly saved up to a week's rental of the video equipment. A comfortable viewing room with lighting-control similar to a small movie-screening room is ideal, and the recorded activity to be evaluated should be displayed on a high-quality, large-screen TV monitor.

Summary of Videotaping Technique

In summary, videotaping, an increasingly feasible data collection technique, has advantages far outweighing its disadvantages in field evaluation of several participants simultaneously engaged in a tactical team task. The enormous data recording capacity, full or dual pictures, and a multi-track sound, accommodate not only the audio-visual record of personnel performances of all scheduled and unscheduled activities, but also all verbal commentary by on-site observers and coordinated display of timer, counter, or other indicator readings when needed as aids to evaluation.

B. Observer Data Sheets

Some form of task flow check sheet is necessary to test design, and undoubtedly an asset to the preparation for and monitoring of any field activity. The Data Sheets used at the Ft. Bragg exercise were designed in three formats for these three observers:

- Test Umpire
- Oral and Visual Observer
- Team Events Observer

Following is a discussion of the advantages and disadvantages encountered in field application of this technique of data collection.

1. Test Umpire Data Sheet

Although the standard operational procedure for the 105 mm howitzer task had been modified somewhat and approved by the battery officers as representing their SOP satisfactorily, field observations proved otherwise. The Umpire's primary task was to verify an acceptable consistency from one team to the next in the sequence of sub-task activities and in the numbers and positions of personnel cooperating in each. The Data Sheet's operational sequence diagram was departed from so frequently in practice that the Umpire could depend upon it very little in fulfilling his responsibility. No attempt was made to keep a complete check-marking record in the columns provided, as had been the intent of the preliminary design. Some comments were written into the column provided for them initially, but this procedure was abandoned in favor of the greater facility of verbal comment into the Umpire's audio recorder, particularly since gusty, cold, and partially rainy conditions prevailed. The Data Sheet served well for pre-test orientation of the Umpire to the task pattern and to his evaluation duties, but had it been used as anticipated during the trials, it would only have interfered with his giving continuous attention to his task of monitoring the rather fast-paced sequence of team coordination.

2. Oral and Visual Observer Data Sheet

The original intent in combining oral and visual into one format was simply to standardize two very similar evaluation sheets and give a single form to both oral and visual observers for monitoring and check-off of scheduled and unscheduled intercommunications. Unfortunately, the telemetry recording commanded the full attention of the intended extra observer and one man was forced to follow both oral and visual signaling. The result of this loading, compounded by the disordered task-flow of the inexperienced teams participating, was the same as with the Umpire: the

value to him of the SOP data listing was marginal. The scrambled sequences of sub-tasks made back and forth page turning necessary and distracted the observer from his direct observation, so that it proceeded in choppy "sampling" views if he attempted to follow the task, enter checks, and make written remarks on-site. While he referred to the sheet, the multi-task team was moving along to a new activity. The practical resolution of this monitoring problem was verbal recording of commentary and team verbal exchanges on the portable audio recorder to the greatest extent possible, in place of written comments and checks on the sheets. The major shortcoming in evaluative worth of the Data Sheets for interactivities proved to be that while the sheet could anticipate only the scheduled verbal or visual signaling, these were relatively easy to monitor and count without an aid sheet compared to unscheduled or repeated signaling. And, unfortunately, the frequency and nature of unscheduled communications among the team proved far more important in statistically differentiating between masked and unmasked performance patterns. The VIO/OIO Data Sheet was later reconstructed, through viewing of the videotapes, into a sequence and detail more representative of the task flow that actually occurred in the field, and the videotape observer used this updated version as an aid.

3. Team Events Observer Data Sheet

This observer, in actual field conditions, made even less use of the Data Sheet technique than the others. His task of assessing the quality of the team and individual performance was not to be based on an arbitrary sequence or description. Therefore, he was forced to abandon the printed version of the task when the actual teams did. Again, however, the sheet was of preliminary orientation value, although minimal in this case because this observer was himself a proficient Reserve Army Officer in charge of similar howitzer training in his Reserve Unit. The broad check-column headings of "General Procedure" and "Deviation" were not detailed enough to aid his evaluation by team and individual errors, unsafe conditions, and overall times and coordination on sub-tasks. These categories had to be specified later on a work sheet used in the actual quantification that took place in videotape viewings.

Summary Data Sheet Evaluation

The Data Sheets, while resulting in some recording of data on-site, were more useful as preparatory design and training aids to the observers. It is questionable whether they should be used at all in the field phase, if another, more adequate technique, such as videorecording, is available to release the human observers for full-time monitoring, free from the need to take notes. In a future Data Sheet design, this study would recommend, as a minimum precaution against the unrealistic layout of the sequence of activity, a preliminary observation of a typical subject team

in action. Much better would be an earlier filming or videorecording of such a typical performance with the task-flow being taken down directly from it for the structuring of the Data Sheets. The resulting design, available well ahead of field activity for orientation should be a much more satisfactory aid to the evaluation team's training. If used in the field, it should be reduced to a single page, or at most a separate page for each discrete sequential sub-task, so that no page turning would be required during actual monitoring.

C. 16 mm Filming

Field recording on film has long been an accepted aid to evaluation of performance. It was used in this study in a fairly routine manner, with the exception that it was only a "back-up" visual record as long as the videotaping proceeded satisfactorily. Nevertheless, several assets of the filming technique over videotaping were used to advantage:

- Mobility of Film Recording
- Color Reproduction from the Field
- Reproduction of Multiple Copies
- Ease of Editing for Presentation
- Additional Perspective on Activity

A brief discussion of these features follows.

1. Mobility of Film Recording

The requirement for "orientation sequences" picturing Ft. Bragg, the general field site and incidental preparations beyond the normal locale of the observation was met through use of the 16 mm camera. Its mobility allowed shots at Ft. Bragg's gate, at the telemetry tent, from the hill overlooking the test site, and at various closeup vantage points on the details of participant performance. Some of these sequences would have been unobtainable with a TV camera because of its electric power requirement and cable-length limits. Most of them would have had to be "staged," losing their "candid" value as visual glimpses at the field operations in process.

2. Color Reproduction

Ability of film recording to retain the colors of field action at only slightly greater cost was an aid to later evaluation via the visual recordings. Color TV for field recording is, of course, well beyond the present state

of the art with anything less than bulky and very costly commercial telecasting equipment. The color and superior resolution of film over TV pictures also improved legibility of number vests on the participants, made details of sub-tasks more understandable, and may possibly have explained detection of a few additional bits of data collected from the films that had gone unnoticed in the videotape viewings. Although certainly not critical to the evaluation, the color was definitely a contribution to later effective presentation of results on a sound film, as required by the study contract.

3. Multiple-Copy Reproduction

Although the single copy of videotaped performance records was adequate for use by a small team of evaluators, any larger distribution would have been difficult because duplication of videotapes is not a widespread commercial process as yet. The processing of additional copies of the filmed record, however, made possible the use of one copy for Edgewood presentation and briefing purposes, another for contractor evaluation and file, and a third, partial version for review by the 82nd Airborne, the cooperating Ft. Bragg agency.

4. Ease of Editing

In routine broadcasting use, videotape can be edited as easily or more easily than film, but the technique and special equipment must be acquired. For the purposes of this study, the filmed record provided a much more satisfactory medium for condensation through editing of specific "clips" into more meaningful brief sequences. For instance, all clearly visible personal errors were clipped out and assembled together to allow comparison for possible common underlying causes related to the masks. Such reorganized sequences were then compiled into a single reel, copied, and available in an unspliced version for repeated presentation. The addition of a magnetic "sound-scribing" strip, rather than a permanent optical sound track, further simplified editing by allowing recording of a commentary on the picture later, in bits and pieces, any portion of which could be erased and modified when later results and recommendations of the study so indicated.

5. Additional Perspective

The film cameras did, of course, offer two additional angles of view in monitoring the activity. This resulted in some additional data collection*, although perhaps not enough to justify the duplicating of many

* For example, a rather serious tumble taken by a member of one masked team was completely undetected on either videotape recording, but clearly pictured on the 16 mm cover shot.

costs of field recording by videorecording alone. A small, but possibly significant, contribution of the additional cameras and cameramen was the effect of almost completely encircling the participants so that it became useless for them to attempt to mask any poor quality or unscheduled behaviors on the side away from the camera or, conversely, to "play to the camera" consciously.

Summary

The more important the presentation, reproduction and dissemination of visual records of a field observation, the greater the justification for filming rather than, or in addition to, videotaping. It is probable, in fact, that for the immediate future, contractual requirements for a visual record will specify film rather than videotape for this reason of universal availability of film presentation equipment. Where evaluation alone is the prime use, field recordings are probably best videotaped, but filmed kinescopes must still be taken of any sections of the tape that are to be distributed to agencies without videotape viewing facilities.

D. Audio Recording Technique

In field evaluation, both participant and observer audio recording is essential. Intercommunication among team members was recorded at Ft. Bragg by the video recording system; but had only silent cameras been used for the visual record, a portable audio recording system would have had to be taken into the field for continuous audio recording during each trial run. A portable magnetic recorder is ideal; it is rugged, requires little power, and no special technical knowledge. Observer audio was recorded both on the extra audio track of the videotape and on individual audio recorders. These small devices proved most useful to the observers for the reasons given below:

1. Rapid-Pace Commentary

Observer rotation of details can progress, via verbal comments into a hand-held microphone, for faster than written comments onto a note pad or even checks on a sheet. Moreover, there is no interruption of visual attention toward the performing team while making such notes. In several instances, audio note-taking was the only possible way of keeping pace with observed activities, communications, deviations, etc.

2. Universal Intelligibility

Although many people develop handwriting styles and symbols so obscure as to be unintelligible to anyone else, their spoken comments are generally understandable to others. Thus, it was possible for a technical assistant, not actually present at the field site, to transcribe the audio

notes of each field observer, and to aid him directly in data reduction. With only written notes, data reduction would have had to wait upon the time and availability of each specific observer for translating his field notes.

3. Aid to Recall

The audio notes, while more easily understood by others, also give the observer himself an excellent frame of reference for recollection of subtle details of the field activity, simply by providing an exact repetition of his phrases, and of the background noises and commands that occurred simultaneously in the field.

4. True Portability

Individual recorders of sufficient quality for voice playback weigh only a few ounces, run for hours on a battery, and will record as many notes as can be conveniently spoken during a normal observation session. Cartridge loading then allows a quick change to a fresh tape until the recorded one can be transcribed and reused. Unless the recorder is pocket-sized or provided with a strap, however, it can interfere with use of the hands. The hand-held microphone should be provided with a lapel clip if the observer needs both hands free at all times.

5. Condensation of Commentary

Since a personal audio recorder may be started and stopped at will, comments may be condensed for later replay in a running time for less than the actual running time of the observed activity. This is an enormous aid to data reduction and analysis. If only a real-time audio track of the activity were available, the analyst would have to monitor extensive periods of silence between pertinent remarks about errors, about communications or whatever. But with personal recorders, an individual observer's remarks concerning errors during an 18-minute run may be replayed, counted, and categorized in two or three minutes.

Personal recorders also eliminate the possible simultaneous recording of two observers' comments on one track and allow individual volume settings for speaking levels soft enough not to distract the performing team.

E. Telemetry Recording Technique

The technical detail of setting up and controlling a heart-rate data telemetry system is a supplementary item; but some remarks might be made on the applicability of this technique in the field. Although the data was collected most efficiently, its value in team performance analysis

was limited. Heart-rate is essentially an individual physiological measurement, possibly better controlled as a meaningful variable in the laboratory where related measures--body temperature, perspiration, respiration, etc.--can be taken simultaneously. Only a portion of the team was monitored. The cost, time and preparations involved in telemetry from the entire team would have been considerable, with the additional possibility of electronic interference problems.

With a small team of two, three, or four men, however, engaged in a physical interactivity in which the timed relationship among their various heart-rates reflected a prime test condition, field use of heart-rate telemetry could have contributed greatly to performance evaluation. At any rate, the electronic data telemetry system itself could be adapted to various collecting tasks other than heart-rate; and the resulting graphic records on time-indexed paper rolls have the same analytic advantages previously cited for the other objective records that may be brought back from the field.

F. Post-Testing Technique

Subjective information from the participants in a performance test must be considered not only necessary to complete data collection, but also of possible high value in detecting unobserved individual errors, serious faults in task design, or extremes in training or attitudes and motivations.

The field environment allows little time and convenience for extensive interview, which is very difficult to quantify and evaluate in any event. The predominantly fill-in and check format used in this study's exercise appears satisfactory for average I. Q. and training levels. More comprehensive briefing and debriefing should be attempted as part of a standard evaluation test only where there is some assurance that the majority of the subjects will be above average in abilities or experience, and that subjective impressions will contribute reliable data unobtainable by the objective recording techniques.

G. Supplementary Photography Technique

The use of a Super 8 motion picture camera and a 35 mm still camera at the site made possible standard, inexpensive photographs of the general activity. Although not needed for formal data collection and evaluation, this photographic technique readily allows production and distribution of short motion sequences and color prints or slides for the effective orientation of personnel, not present on-site, to the actual field environment activity.

H. Summary Recommendation: Techniques for Standard Team Task Evaluation

Each of the performance measures investigated in this study and each of the associated field evaluation techniques just discussed has its particular merits and shortcomings. For a consistent framework in which a given factor, such as the effect of protective clothing, can be evaluated through an extended series of comparative tests, a selection must be made of only the most useful measures and most feasible techniques. The exact combination and the details may vary slightly with the specific object of evaluation, but a recommended procedure and structure for general team evaluation from field performance can be described as a result of the knowledge gained in this study.

This recommended approach overcomes many of the traditional difficulties in monitoring several individuals simultaneously, in quantifying interactivities among them, in obtaining a sufficient "n" of observations under similar conditions, and of doing all this in the field environment for greater validity. In other words, it is an improved approach to the field evaluation of team performance. Briefly stated, these are the recommended procedures:

(1) Videorecord a "typical performance" of the task to be evaluated, with a descriptive commentary track by someone knowledgeable in the task, some time ahead of field observations.

(2) View this recording repeatedly to detail the operational sequence of team interactivities required, divide the task into functional sub-tasks, and generate a task flow chart.

(3) Compile a "script" of all required or "scheduled" team actions under oral and visual intercommunications, and under required and critical team coordinations, flagging their points of probable occurrence in each sub-task along the functional flow chart as an observer aid.

(4) Schedule repeated field observations of as large a number of teams as feasible, using counterbalancing sequence of control and experimental conditions. (Trials should also be included in which some teams perform only with masks twice or more in succession, and others perform only without masks twice or more in succession, in order to better isolate the sequence effect observed in this initial study).

(5) Monitor these field trials with videorecording from at least two camera angles (detailed and wide) and at least two human observers, one "expert" in the task and one monitor-in-charge to direct the

cameras through the critical task sequences. If economically feasible, three additional human observers should be on-site, each equipped with a portable audio recorder for commentary: observers of oral, visual, and error aspects of the team's performance, trained by earlier viewings of the videotape of typical performance.

(6) Review the videotaped performances in a series of "passes" scheduled in deliberately reordered sequences to counterbalance observer training effect. Each pass should be used to extract a separate variable in depth by specified observing assignments. (Times, errors, counts of communications, etc. are all quantified in this phase in the manner described in Section V of the report).

(7) Reduce the data by variables, sub-task, and sequence, as described in Section IV of this report, to produce totals for generation of standard graphs of the three types used in this report: task profiles by individuals, condition and sub-task (Section IV, Figures 12-16); variable bar graphs by condition and sub-task (Appendix C) and sequence/mask effects separation graphs (Appendix B).

(8) Maintain this standard procedure for the same field task and the same observed variables over a sufficient "N" of teams to produce a stabilising of the resulting deviations of variables within a reliable area; i.e. to produce a "data base" for establishing statistical significance of the differences observed in performance with protective equipment.

Evaluation of Appropriateness of Tasks

Whether structuring a field task into a standard test or evaluating a task already advanced as a tentative test, an analysis can be applied using the experience of this study to adjudge the sensitivity of the task to the desired team performance variables--and to modify it if necessary, for improved generation and collection of these data.

In these howitzer exercises, for example, later analysis revealed most trends from the data in sub-tasks two, three and five (Section IV, Table H). Sub-tasks one and four proved too short, repetitive, routine or otherwise unsuitable in providing the most valuable data. In order to evaluate a task for probable high yield of data in the areas of team performance of most interest, the following questions should be asked:

Are these sub-tasks appropriate for intercommunications?

Since this area of variables yielded significant differences with masks even with the small "n" of this study, analysis of sub-tasks of a

tentative test should give it priority attention. At least one sub-task should contain a large number of required oral interactivity and a second one a large number of required visual interactivity. These should be separate sub-tasks ideally so that any observed decrease in scheduled oral signaling is not offset, in the total communications data count, by a compensating increase in visual signaling, and vice versa. A sub-task in which every man works always on his own will require virtually no intercommunication, regardless of protective equipment, and a sub-task in which everyone works together in physical contact and practical unison will likewise require little oral or visual signaling in either control or experimental condition. The sub-tasks appropriate as team communications measures, therefore, will possess a pattern of necessary coordination among the individual personnel by a series of signals. For the visual sub-task, team members should be working several feet or yards apart in full view of each other and at duties requiring interactivity. For the oral sub-task, team members should be working largely within sound but not within sight of each other. Since the change in ratio of scheduled to unscheduled oral communication seems a particularly sensitive variable in the communications data, an ideal sub-task should be one in which all team members, not just the group leader, are permitted (though not necessarily required) to speak at various points for the proper coordination of the field activity.

The inclusion of sub-tasks for oral and visual team intercommunication, assessed by the above criteria, should assure an effective generation and collection of this set of performance variables in a standard fiel test.

Is there a sub-task appropriate for error differences?

If the activity is so routine that a "crack team" may become highly practiced in its unchanging performance, then few errors will occur under any circumstances. If the activity requires so much stamina, learning, or individual adaptability that few can fully meet the demands, then errors will occur regularly even under optimum conditions. In this study sub-task two, laying the howitzer, required more training and individual initiative than the inexperienced troops generally possessed; the result was a relatively high error rate in both conditions. Sub-task four, repeated firing for effect, involved only half the team in critical and repetitive routines; the result was an extremely low rate of error in both conditions. A sub-task appropriately sensitive to error differences should be of moderate difficulty for average troops, require a variety of action, and extend long enough in time to test their ability to sustain a careful performance beyond critical moments and into fatigue. Sub-task five, march order, though a bit short, was generally appropriate for error sensitivity because it involved everyone, required some initiative, was considered less tactically critical by the participants, and occurred at the time of greatest probable fatigue. Because it required little coordinated activity among two or more men, sub-task five

was not as ideal for generating team errors as for personal error data.

Is there a sub-task appropriate for time differences?

Time data must be taken in depth if it is to be a useful measure differentiating performance of teams in protective clothing. A real task always has a certain amount of "slack time" for the various members, particularly in the field under training rather than tactical conditions. This slack may become a buffer that simply shrinks a bit during masked performance with the total task time remaining approximately the same. Since sequence, or practice effect, is far more likely to affect total time, only a careful attention to the pattern of the time variable can render it useful. This is the prime purpose in this study of the density profiles (Section IV, Figures 12-16), the division of time into sub-task totals (Appendix C, Figure 11), and the separation of sequence from mask time effects (Appendix B, Figure 15). Thus, although equal numbers of teams in this study performed faster with and without the masks, this apparent equality was demonstrated as due to sequence counterbalancing. Analysis in depth showed slower times with masks for each and every sub-task and one improvement in the time of second trial without mask three times as great as another when the second trial was with masks.

The sub-task most appropriate for time data seems to be not so short as to cause observer lag or error in time recording to become a major factor (sub-task one averaged only 45 seconds) and not so long as to allow the team's "slack time" to become a dominant factor (in sub-task three half of the team did nothing unless the other half needed help). A moderately long sub-task is also more sensitive if several concurrent small group or individual activities must be terminated in coordination before the task can move along; a delay in a one-man activity is thus magnified into a delay of the entire team time at several points in the task flow if the experimental condition is, in fact, troublesome. This was the case in sub-task two, which proved most sensitive to the time variable at Ft. Bragg. If the sub-task flow analysis reveals a lack of interconnection among simultaneous individual or two-man tasks, and no preferred sequence of their performance, then the total time is less likely to suffer from accumulated effects of small delays or individual slowness. For good time data, therefore, at least one sub-task of moderate duration should require all members to work a major portion of the total time at a logical sequence of inter-related individual or small group tasks.

Summary of Recommendations on Evaluation Techniques

By using the list of recommended field evaluation procedures given above, and including the effective measures of team communications, errors, and performance times described in the preceding paragraphs, a

protective equipment evaluation program of field testing should readily acquire a standard format and a reliable data base.

This effort should be greatly facilitated by modeling to standard usage the data collection and analysis techniques developed in this study with emphasis on standard data display formats (Section IV) and videotape evaluation (Section V).

SECTION VI

SUMMARY

A. Conclusions from Current Study

1. From Field Data

All six hypotheses put forth in initial design were substantiated, in some degree, in the field observations. Within the limited number of runs, statistical significance was achieved in only two instances, but no tendency of the data was contrary to the direction hypothesized.

Briefly, these were the anticipated and actual effects of the mask on team performance in the field:

Total time was predicted to be slower with masks and, in fact, was slower in each of the five sub-tasks. The total task took the teams a mean time of 9.6 minutes without the masks and 11.0 minutes when wearing them (Appendix C, Figure 11).

Total errors were predicted to increase with masks and did, in fact, with particular effect on personal errors in the final sub-task. (Section IV, Figures 9, 10, and 11).

Overall quality of performance was predicted to be poorer with masks. From the participants' subjective views this was true, both of their team and of themselves individually--of those assessing test performance as worse than normal, twice as many did so after a masked run (Section IV, Table E). From the "expert" observer's view, no discernable difference could be detected, although the bare discrepancy in magnitude of his data was in the predicted direction. (Appendix C, Figure 14).

Density of activity was predicted to be greater with masks. For the central and tactically critical sub-task three, this was true for seven of the eight members of the team. For the less practiced preparation and departure activities of sub-tasks two and five, however, an exactly opposite pattern emerged, so that the total task effect was a near-cancelling of all differences in the density data (Section IV, Figures 12-16).

Visual interactivity was predicted to increase among masked team members, which it did, significantly, in field testing. Unscheduled manual signaling for visual coordination of the laying activities of sub-task two was greater, in particular (Appendix C, Figures 2, 4, 6, and 8).

Oral interactivity was predicted to decrease with masks, at least for "effective" communications, although the hypothesis noted that "masking might produce more frequent oral attempts at coordination." This, in essence, was the actual case in the field: "scheduled" orals decreased slightly with masking, but unscheduled and repeated orals increased (Appendix C, Figures 1, 3, 5, and 7).

A corollary to the oral hypothesis, predicting fewer signals with hoods, could not be tested since only two teams performed in that condition. However, the increase in oral signaling in both instances, even outweighing the usually dominant effect of sequence in progressively reducing oral interactivity (Appendix B, Figure 20), indicates that the corollary might be false. This would possibly be because an essential minimum of oral/aural interchange exists in coordinating the task, and when attempted through the combined mask/hood barrier to speech and hearing, is augmented by more frequent attempts and repetitions to assure the complete communication of these essentials.

2. From Overall Study

Several general conclusions about effective evaluation of field teams have been drawn from this research, and are incorporated in recommendations (Section V, H). In summary, these are the principal points:

Analysis and design effort prior to field observation is the most effective means of establishing adequate experimental control in the field environment. The task to be observed must be divided into sub-tasks and defined in detail for the intended observers, ideally via a visual recording of a typical team's performance. Observers should then be trained in non-overlapping duties and supervised on-site by a monitor-in-charge, who is himself free of any data collecting responsibilities during the field exercises.

Videotaping was found the most advantageous means of data recording in the field, particularly when supplemented by audio recording of the teams' verbal exchanges and the observers' commentary directly on separate tracks of the videotape. Individual portable audio recorders were recommended for observer note-taking as being more efficient and less distracting than manual check sheets or clipboards for notes. The telemetry of physiological data, unless basic to the specific evaluation, was determined to be more appropriate in the laboratory environment where it can be integrated with other individual data collectors. Post-performance questioning of the participants was found advisable as a subjective source of evaluative data. Usefulness of observers on-site was found improved by making greater use of their judgment and interpretive abilities and relegating their more straightforward timing and counting duties to the videorecorders for later quantification during scheduled viewings of the tapes when the field activities are completed--either by the original observers or by objective analysts.

Standard presentation formats were recommended as aids to reduction and analysis of the collected performance data. For defining the nature and general pattern of a task, a bar profile chart was developed that displays the relative activity of various individuals within a team during each sub-task and the changes in these proportionate duties from one sub-task to the next (Section IV, Figures 12-16). For finding the sub-task most sensitive to each variable a five-fingered bar graph was developed (Appendix C) that could be applied to any field task in conjunction with discriminability formulae described in Section IV. For separating mask effect from sequence effect for each variable, a line graph was developed (Appendix B).

Evaluation techniques were outlined for data generation and interpretation from field team tasks in general. Methods both for estimating and statistically testing for discriminability of a variable were described, and it was concluded that a structured field test can probably dispense with those sub-tasks showing least discriminability for all important variables. In the 105 mm howitzer task of this study, for example, virtually all significant differences and trends were derived from sub-tasks two, three, and five (Section IV, Table H). Sub-tasks one and four added little to the results, but much to the time of the exercise (and to the cost, since sub-task four included the firing of two rounds of live ammunition for effect). Recommended procedures, based on the experience gained in this study, were listed in Section V, H, for effectively conducting field observations of teams, collecting the data, and analyzing the results against the standard presentational coordinates provided for initiating a data base.

The findings and recommendations of this study are believed to be a sufficient basis for the establishment of a workable standard procedure for evaluating the effects of various protective equipments developed at Edgewood Arsenal Research Laboratories on the tactical performance of teams in the field.

3. Summary of Conclusions

With the details of this report always borne in mind, the following generalizations may be advanced as a result of the study:

a. A simple measurement design is inadequate for field testing of teams. Observations must be structured to detect separate patterns of performance across multiple variables, individual team members, and various sub-tasks.

b. Human observers at the field site cannot alone record this wide range of data at the pace of team performance. Data recording equipment, particularly videorecorders, assure a more adequate and objective data collection.

c. Performance of certain team tasks, generally those requiring individual initiative in coordinating activity, may be significantly altered by the wearing of protective masks. Other tasks of a repetitive drill nature may be much less affected. Even very meticulous testing during field training activity may fail (because of the "slack time" and indirect motivations inherent in the practice environment) to reflect validly potential decrements in a tactical environment.

d. For 105 mm howitzer crews specifically, there will probably be no consistently critical performance problems while wearing protective masks, although the team's normal pattern of oral intercommunication is definitely disrupted.

B. Recommendations for Further Study

The results of this effort offer three promising directions for further study in the areas of field collection equipment, team data analysis, and standard performance testing.

1. A system of team data reduction and analysis in depth should be sophisticated, so that large quantities of data from multiple trials over time could be methodically built into a data base for each variable, team member, and sub-task of a standard test. Application of automatic data processing and mathematical modeling should be investigated; and weak spots in the initial design, such as the quality and density of activity variables should be strengthened through quality control and systems analysis techniques.

2. The generalizability of objective field monitoring techniques, such as videorecording should be explored in detail. Any field performance that is broadly repeatable could become a standard test, in effect, by visually recording each trial and redirecting the structuring control from the participants to the observation procedure: standardly structured camera coverage, videotape editing and playback, observer training, and video viewing schedules by separate variables and sub-tasks.

3. Routine team testing should be conducted using the improved data collection devices and data analysis techniques perfected in the two studies described above. The current study emphasized the detection of pattern changes, not the significance or tactical validity of all such changes. In its field observations, this objective was achieved. The promising trends in variables from this small number of trials should now be pursued in a much larger "n" to provide the data base for probable statistical significance in several categories. These extended performance tests, if done with chemical protective equipment, could then possibly establish reliable patterns of difference valuable in improving design of the equipment and in predicting masked team behavior under valid tactical conditions.

APPENDIX A

BIBLIOGRAPHY

All literature surveyed and found pertinent to this study is listed alphabetically in the following bibliography, which covers three general areas of research:

1) General Test Design
2) Protective Equipment
3) Team Performance

American Institute for Research (AIR), Procedures for Evaluating Research Personnel with a Performance Record of Critical Incidents, for Office of Naval Research, June 1950.

Andersson, Bengt-Erik & Nilsson, Stig-Goran. Studies in the Reliability and Validity of the Critical Incident Technique, J. of Appl. Psychol., 1964, 48, No. 6.

Babs, R.F. A Set of Categories for the Analysis of Small Group Interaction, Amer. Society Review, 1950, 15, 257-263.

Baker, R.A. & Warnick, W.L. Determination of Combat Job Requirements for Armored Cavalry Platoon Personnel, HumRRO Tech. Report 92, Ft. Knox, Ky., Dec. 1964.

Bartley, S.H. & Chute, Eloise. Fatigue and Impairment in Man, McGraw-Hill Book Co., New York, 1947.

Bavelas, A. Communication Patterns in Task-Oriented Groups, J. Acoustical Soc. Amer., 1950, 22, 725-730.

Beecher, N. U.S. Might Step Up the Use of Non-Lethal Gas in the War in Vietnam, The Wall Street Journal, New York, Jan. 5, 1966.

Beecher, W. Chemicals vs. the Viet Cong--"Right," or "Wrong?" The National Guardsman, Feb. 1966.

Bendig, A. W. Reliability and the Number of Rating Scale Categories, J. Appl. Psychol., 1954, 38, 38-40.

Bittner, R. & Lindquist, E.A. The Rank Comparison Rating Method, J. Appl. Psychol., 1950, 34, 171-177.

Combat Div. Command. Road Battalion Operations in a Toxic Environment, Report 63-4, Headquarters, USACDEC, Ft. Ord, Calif., Dec. 1963.

Cronbach, L.J. Test "Reliability": Its Meaning and Determination, Psychometrika, 1947, 12, 1.

Cummings, E.G., et al. The Effect of Protective Masks on the Soldier's Ability to Run the Half-Mile. CWLR 2254, 1958.

Davies, C.N. Design and Use of Respirators, New York: Pergamon Press, MacMillan Co., 1962.

Dunlap and Associates, Inc. Development of Methodology for Measuring Effects of Personal Clothing and Equipment on Combat Effectiveness of Individual Soldiers, Prepared for U. S. Army General Equipment Test Activity, Ft. Lee, Va., Phase I Report, Feb. 1964, Phase II Report, Dec. 1964.

Dunlap and Associates, Inc. Feasibility Study, Army Study Requirement No. 7, CBR Protection of Combat Troops, Proposal 60-21, Sept. 12, 1960.

Dunlap and Associates, Inc. Protective Mask Configuration Concept Development, Proposal 63-49 submitted to U.S. Army Chemical Center Procurement Agency, Edgewood Arsenal, Md., May 1963.

Eckenrode, R.T. Operational Problem Solving and Decision Making, Dunlap and Associates, Inc., Darien, Conn. Unpub. 1961.

Electronic Device Measures Fatigue, San Diego, Calif.: United Press International, Sept. 28, 1965.

Ely, J.H. Data Collection for the Design and Evaluation of Man-Machine Systems, ASME Publ., 1958.

Farber, C.F. Something Different About Little John, Artillery Trends, U. S. Army Artillery and Missile School, Ft. Sill, Okla., April 1965.

Finan, J.L., Sarah C. & Hartson, L.D. A Review of Representative Tests Used for the Quantitative Measurements of Behavior-Decrement Under Conditions Related to Aircraft Flight, USAF Tech. Report No. 5830, United States Air Force, Air Material Command, Wright-Patterson AFB, Ohio, July 1949.

Floyd, W.F. & Welford, A.T. (Eds.) Symposium on Fatigue, London: H. K. Lewis & Co., Ltd., 1953.

Fogel, L.J. Biotechnology: Concepts and Applications, New Jersey: Prentice Hall, Inc., 1963.

George, C.E., Hoak, G R. & Boutwell, J. Pilot Studies of Team Effectiveness, U.S. Army Infantry Human Research Unit, Ft. Benning, Ga., Feb. 1963.

Goldberg, M. N., et al. Individual Respiratory Protection Against Chemical and Biological Agents, Prepared for U. S. Army Chemical Research and Development Labs, Edgewood Arsenal, Md., Phase I Report, October 1965, Phase II Report, North American Aviation, Inc., January 1965.

Grace, G. L. Application of Modern Technology to Discovering and Measuring the Abilities of Man, SP-1645, System Development Corp., Santa Monica, Calif., July 1964.

Granzeier, F. J. Toxic Weapons, Indus. Research, Aug. 1965, 69-74.

Gruber, A., Dunlap, J. W. & DeNittis, G. Development of Methodology for Measuring Effects of Personal Clothing and Equipment on Combat Effectiveness of Individual Soldiers, prepared for U. S. Army General Equipment Test Activity, Ft. Lee, Va., Dunlap and Assoc., Inc., Darien, Connecticut, Dec. 1964.

Gulliksen, H. Paired Comparisons and the Logic of Measurement, Psychol. Rev., 1946, 54, 199-213.

Hays, W. L. Statistics for Psychologists, New York: Holt, Rinehart, and Winston, 1963.

Headquarters, 2nd Armored Division. After Action Report Exercise "Retort," Ft. Hood, Texas, Nov. 1964, 23-25.

Hoisman, A. J. & Daitch, A. M. Techniques for Relating Personnel Performance to System Effectiveness Criteria: A Critical Review of the Literature, Dunlap and Associates, Inc., Santa Monica, Calif., Sept. 1964.

Horsfall, A. B. & Arensberg, C. M. Teamwork and Productivity in a Shoe Factory, Some Theories of Organization, Rubenstein and Haberstroh, Dorsey Press, Homewood, Ill., 1960.

Kurke, M. I. Operational Sequence Diagrams in System Design, Human Factors, 1961, 3, 66-73.

Lanzetta, J. T. & Toby, T. B. Group Learning and Communication as a Function of Task and Structure Demands, J. Abnorm. Soc. Psychol., 1957, 55, 121-131.

Lind, A. R., et al. Reactions of Mines Rescue Personnel to Work in Hot Environments, Medical Research Memorandum 1, National Coal Board, London, 1965.

McCormick, E. J & Backus, J. F. Paired Comparisons Ratings: I. The Effect on Ratings of Reductions in Numbers of Pairs, J. Appl. Psychol., 1952, 36, 123-127.

McCormick, E. J. & Roberts, W. K. Paired Comparison Ratings: II. Reliability of Ratings Based on Partial Pairings, J. Appl. Psychol., 1952, 36, 188-192.

McGrath, J. E. A Summary of Small Group Research Studies, AFOSR Document No. 2709, Human Sciences Research, Inc., Arlington, Va., June 1962.

McNemar, Q. Psychological Statistics, New York: John Wiley & Sons, Inc., 1962.

Madden, J. M. A Comparison of Three Methods of Rating Scale Construction, J. Industrial Psychol., 1964, 2, 43-50.

Marks, M. R. Development of Human Proficiency and Performance Measures for Weapon Systems Testing, ASD Technical Report, 61-733, Behavioral Sciences Lab., Aerospace Med. Research Labs., Aeronautical Systems Div., Air Force Systems Command, Wright-Patterson AFB, March 1962.

Meister, D. Methods of Predicting Human Reliability in Man-Machine Systems. Human Factors, 1964, 6.

Meister, D. & Rabideau, G. F. Human Factors Evaluation in System Design, New York: John Wiley & Sons, Inc., 1965.

Patton, R. M. Behavioral Testing During a Seven-Day Confinement, The Information Processing Task, Tech. Note D-1973, National Aeronautics and Space Administration, Washington, D.C., December 1963.

Rabideau, G. J. Field Measurement of Human Performance in Man-Machine Systems, Human Factors, 1964, 6, 663.

Rabideau, G. J. Problems of Measurement Involved in Field Testing, Report No. GM 6300 5-1593, Space Tech. Labs.

Rapoport, A. In Search of Quantifiable Parameters of Group Performance. In Systems: Research and Design, Wiley and Sons, Inc., New York 1961.

Rappoport, M. Further Consideration about the What and How of Field Testing, Paper presented at the Western Psychol. Assoc., Seattle, Wash., June 1962.

Roby, T.B. Prerequisites for Pair-Scores to be Used for Assembling Small Work Groups, Research Bulletin AFPTRC-TR-54-13, Air Force Personnel and Training Research Center, Lackland AFB, Texas, 1954.

Roby, T. & Lanzetta, J. Considerations in the Analysis of Group Tasks, Psychol. Bull., 1958, 55, 88-101.

Posner, M.I. Information Reduction in the Analysis of Sequential Tasks, Psychol. Rev., 1964, 71, 491-504.

Safety Plan and Scenario, First Armored Division Chemical Field Training Exercise, 16-21, Nov., 1964.

Schutz, W.C. Construction of High Productivity Groups, Studies in Group Behavior, Tufts College, Medford, Mass., 1953.

Sells, S.B. (Ed.) Symposium on Pattern Analysis, USAF School of Aviation Medicine, Randolph AFB, Texas, 1955.

Selye, H. & Fontier, C. Adaptive Reaction to Stress, Psychosomatic Medicine, 1950, 13, 149-157.

Siegel, A.I., Lanterman, R.S. & Wolf, J.J. Techniques for Evaluating Operator Loading in Man-Machine Systems, Contract Nonr-2492 (00), Applied Psychol. Services, 1963.

Shephard, R.J. Ergonomics of the Respirator, Design and Use of Respirators, C.N. Davies, (Ed.), McMillan, Inc., New York, 1962.

Smode, A.F., Gruber, A. & Ely, J.H. The Measurement of Advanced Flight Vehicle Crew Proficiency in Synthetic Ground Environments, MRL-TDR-62-?, Behavioral Sciences Lab., 6570th Aerospace Med. Research Labs., Aerospace Med. Div., Air Force Systems Command, Wright-Patterson AFB, Ohio, Project No. 1710, Task No. 171003, Dunlap and Associates, Inc., Darien, Conn., Feb. 1962.

Smode, A.F. & Yarnold, K.W. Recent Innovations in Methodology for Training and Training Research, Prepared for Office of Naval Research, Washington, D.C., Dunlap and Assoc., Inc., March 1960.

Spacelabs, Inc., A Dynamic Physiological State Monitor, Prop. 9953, Van Nuys, Calif., April 1965.

Swain, A.D. Some Problems in the Measurement of Human Performance in Man-Machine Systems, Human Factors, 1964, 6, 687-701.

Swain, A. D., Altman, J. W & Rook, L. W. Human Error Quantification, Report No. SCR-610, Sandia Corp., Albuquerque, New Mexico, April 1963.

Thibaut, J. W. & Kelley, H. H. Interference and Facilitation in Interaction, the Social Psychology of Groups, Wiley and Sons, New York, 1959.

Torrance, E. P. Crew Performance in a Test Situation as a Predictor of Field and Combat Performance, Human Factors Operations Research Labs, No. 33, Air Research and Development Command, Bolling AFB, Washington, D.C., March 1953.

U. S. Army Chemical Research and Development Labs (CRDL), Seventy-two Hour Outdoor Wearing Trial of Individual CBR Protective Clothing, Tech. Rept. 3097, Army Chemical Center, Md., Sept. 1961.

U. S. Army Field Manual, Chemical, Biological and Radiological (CBR) and Nuclear Defense Training Exercises, FM 21-48, Sept. 1964.

U. S. Army Field Manual, Civil Disturbances and Disasters, FM 19-15, December 1964.

U. S. Army Field Manual, Soldier's Handbook for Chemical and Biological Operations and Nuclear Warfare, FM 21-41, April 1963.

U. S. Army Field Manual, Tactics and Techniques of Chemical, Biological and Radiological (CBR) Warfare, FM 3-5, Nov. 1958.

U. S. Army Pictorial Service. Combat Effectiveness Test Course, Phase II, Ft. Lee, Va., 1965.

U. S. Army Pictorial Service. Report of Field Tests with Psychochemical Compounds. (Confidential) M.'-3-9121. Ft Bragg. N.C., 1959.

Use of Gas in Vietnam Draws Protest. Business Week. March 1965, 34.

Viteles, M.S. & Thompson, A. S. The Use of Standard Flights and Motion Photography in the Analysis of Aircraft Pilot Performance, Report No. 15, Civil Aeronautics Administration, Div. of Research, Washington, D.C., May 1943.

Wagner, R. F. Using Critical Incidents to Determine Selection Test Weights. Personnel Psychol., 1951, 4.

Warnich, W. L. & Baker, R. A. Determination of Combat Job Requirements for Armored Cavalry Platoon Personnel. Report 92, U.S. Army Human Research Unit, Ft. Knox, Ky., Dec. 1964.

Wulfeck, J.W., Weisz, A., and Raben, M.W. Factors Influencing Vision Outside the Aircraft, Chapter IX in Vision in Military Aviation, 1958, WADC Tech. Rept. 58-399.

Zajonc, R.B. The Requirements and Design of a Standard Group Task, J. of Exper. Soc. Psychol., 1965, 1, 71-88.

Zajonc, R.B. Social Facilitation, Science, 1965, 149, 269-274.

APPENDIX B

SEQUENCE DATA

The necessity for separately displaying sequence and mask effects was explained in Section IV. B. 1. c. of this report. The technique developed for such presentation is applied for each variable of this study to generate the standard graphs shown on the following pages.

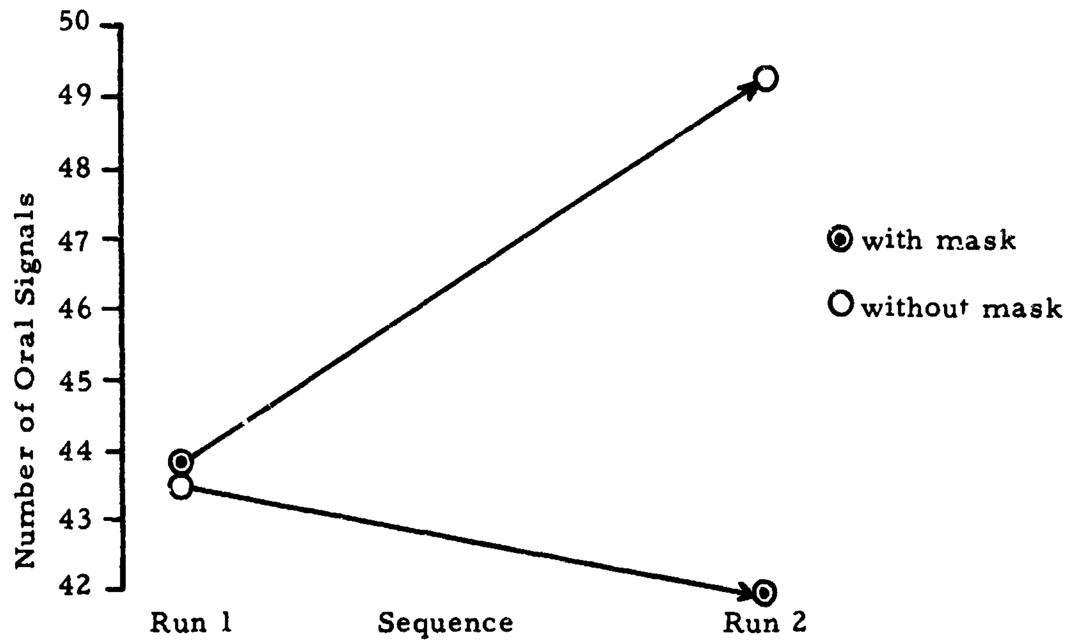


Figure 1 . Profile Lines of Scheduled Orals per Task.

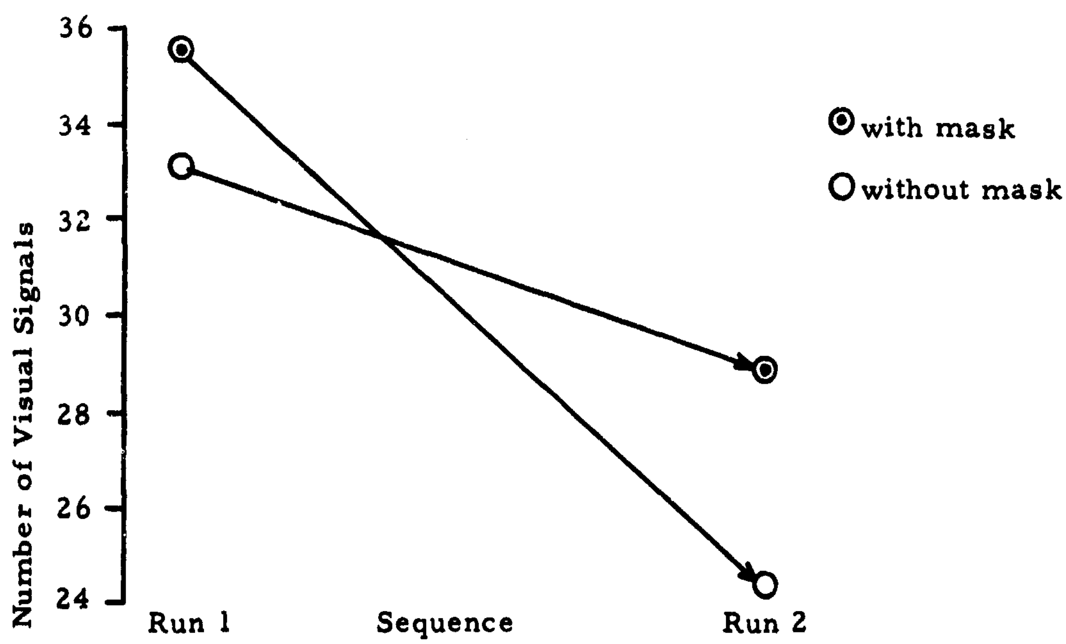


Figure 2 . Profile Lines of Scheduled Visuals per Task.

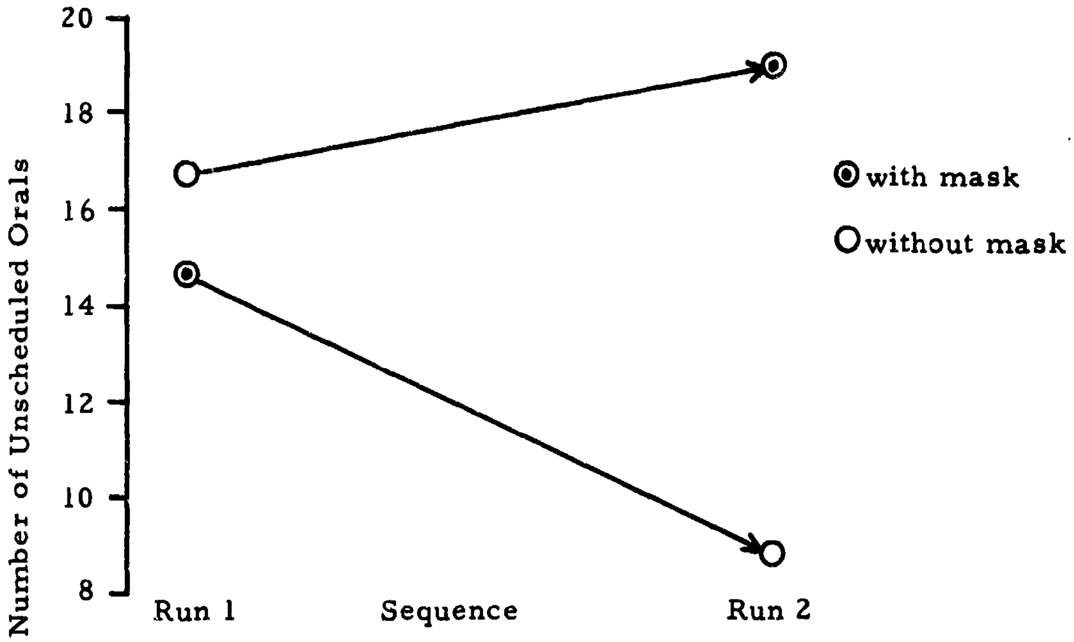


Figure 3 . Profile Lines of Unscheduled Orals per Task.

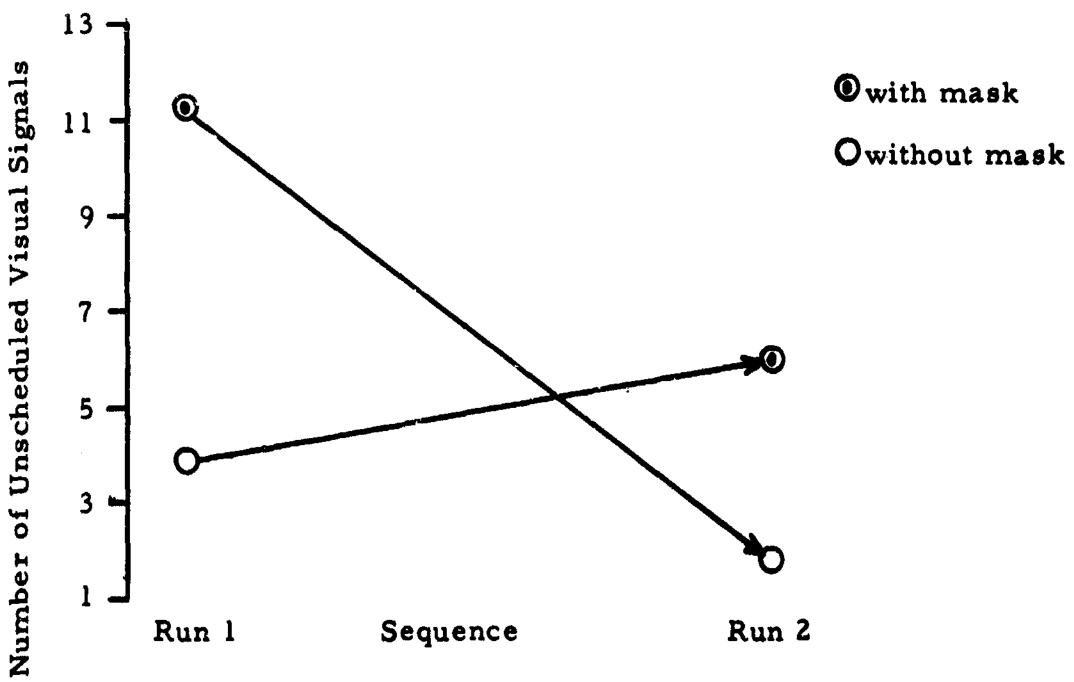


Figure 4 . Profile Lines of Unscheduled Visual Signals per Task.

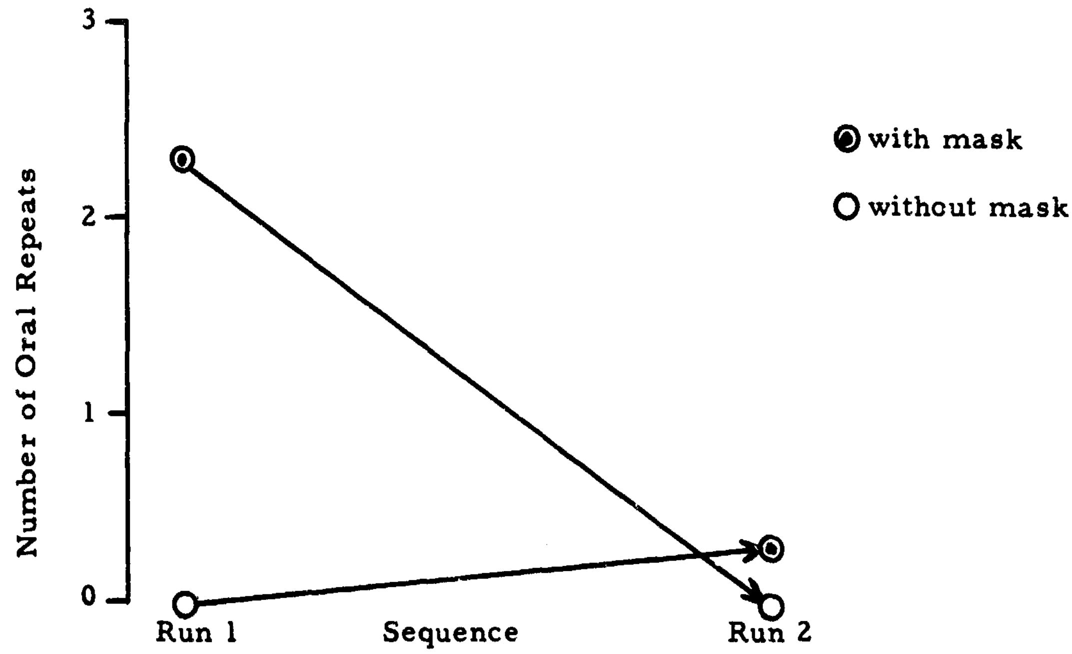


Figure 5. Profile Lines of Oral Repeats per Task.

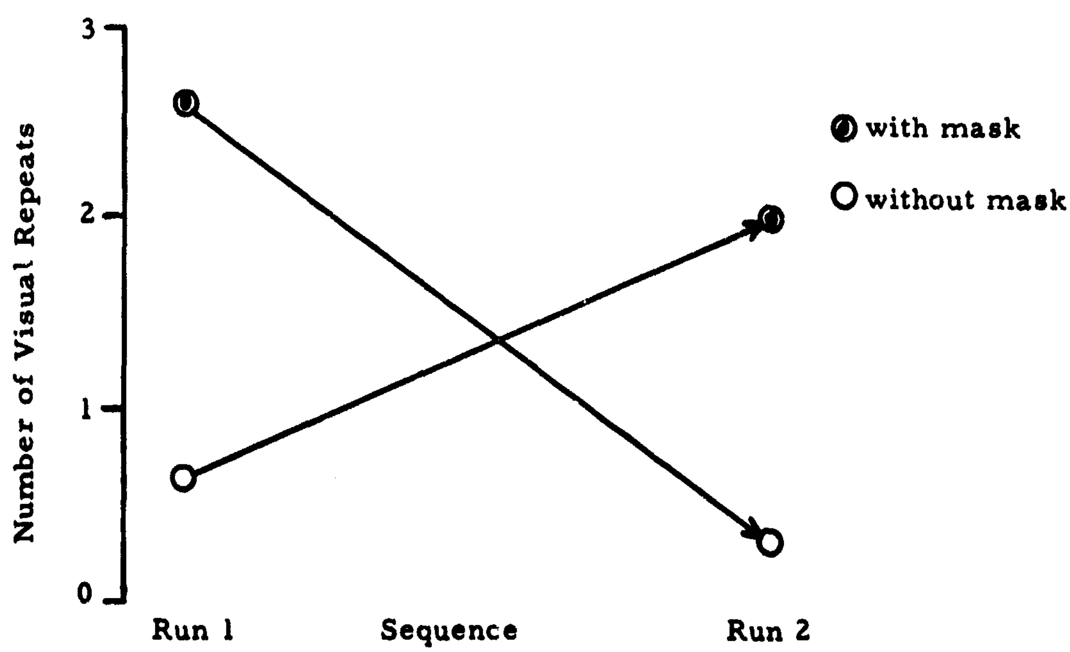


Figure 6. Profile Lines of Visual Repeats per Task.

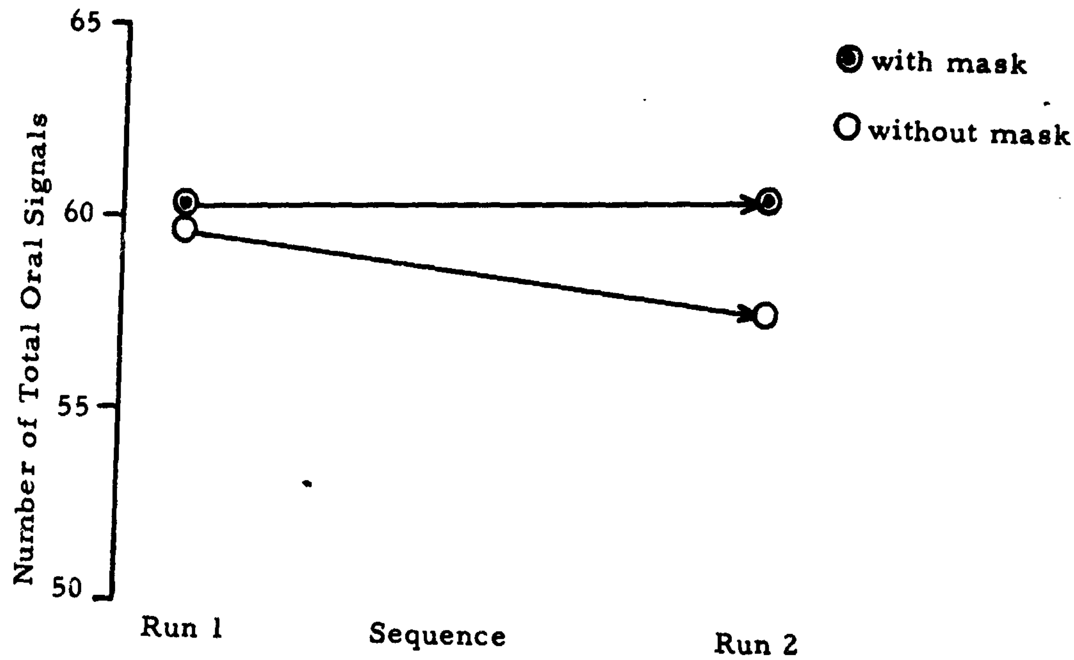


Figure 7. Profile Lines of Total Oral Signals per Task (Scheduled Orals, Unscheduled Orals, and Repeats).

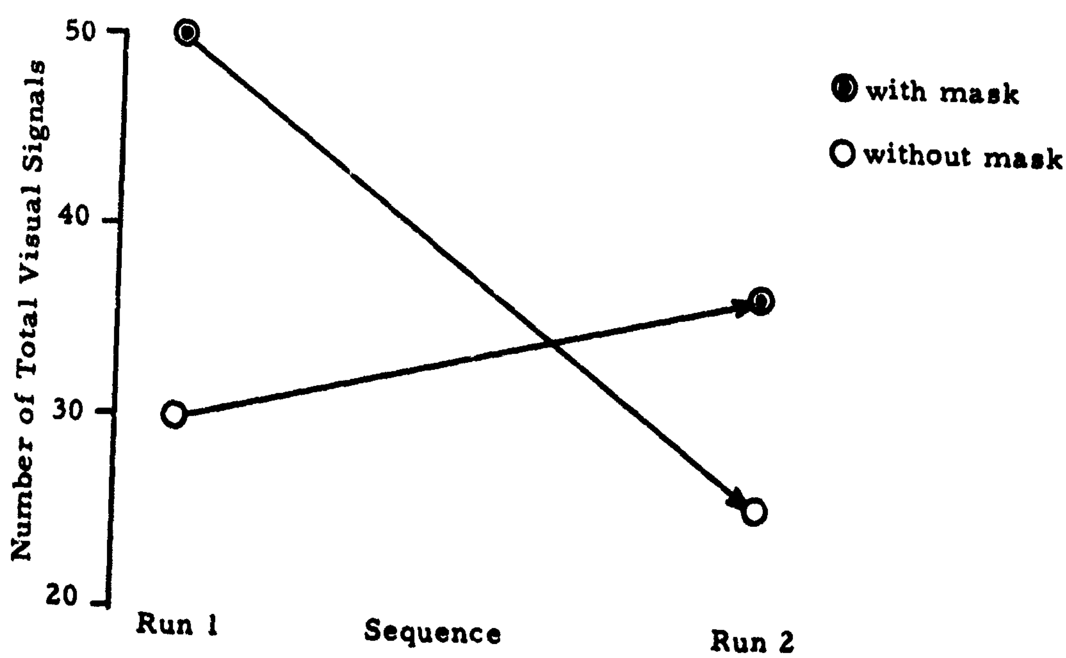


Figure 8. Profile Lines of Total Visual Signals per Task (Scheduled Orals, Unscheduled Orals, and Repeats).

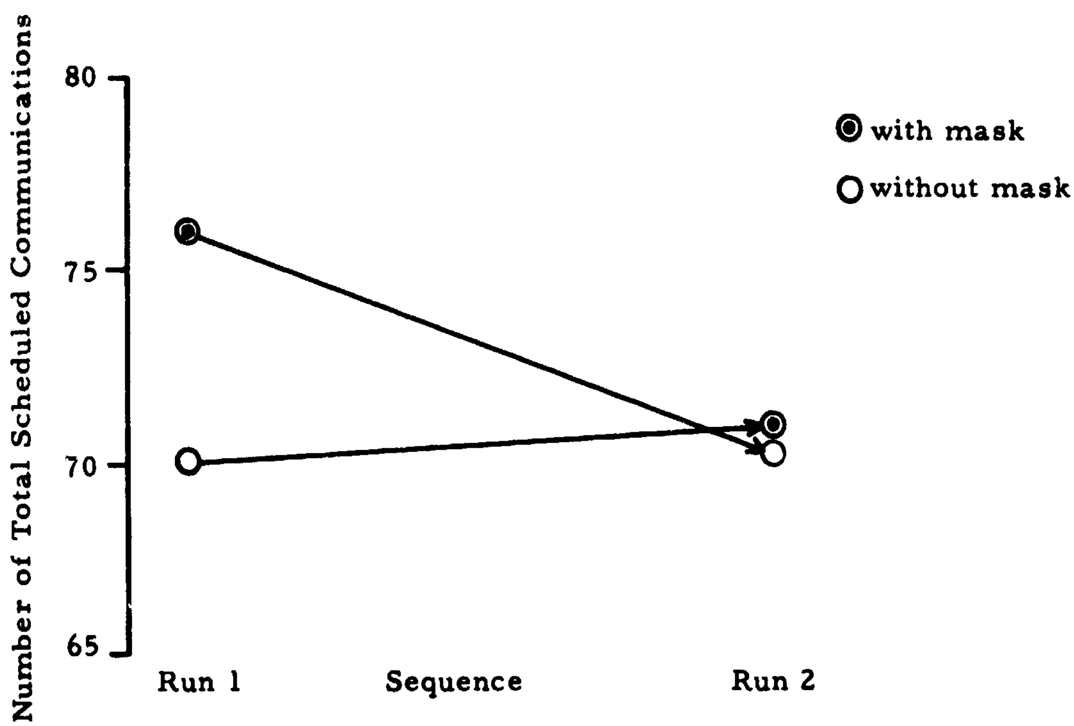


Figure 9 . Profile Lines of Total Scheduled Communications (Scheduled Orals and Scheduled Visuals).

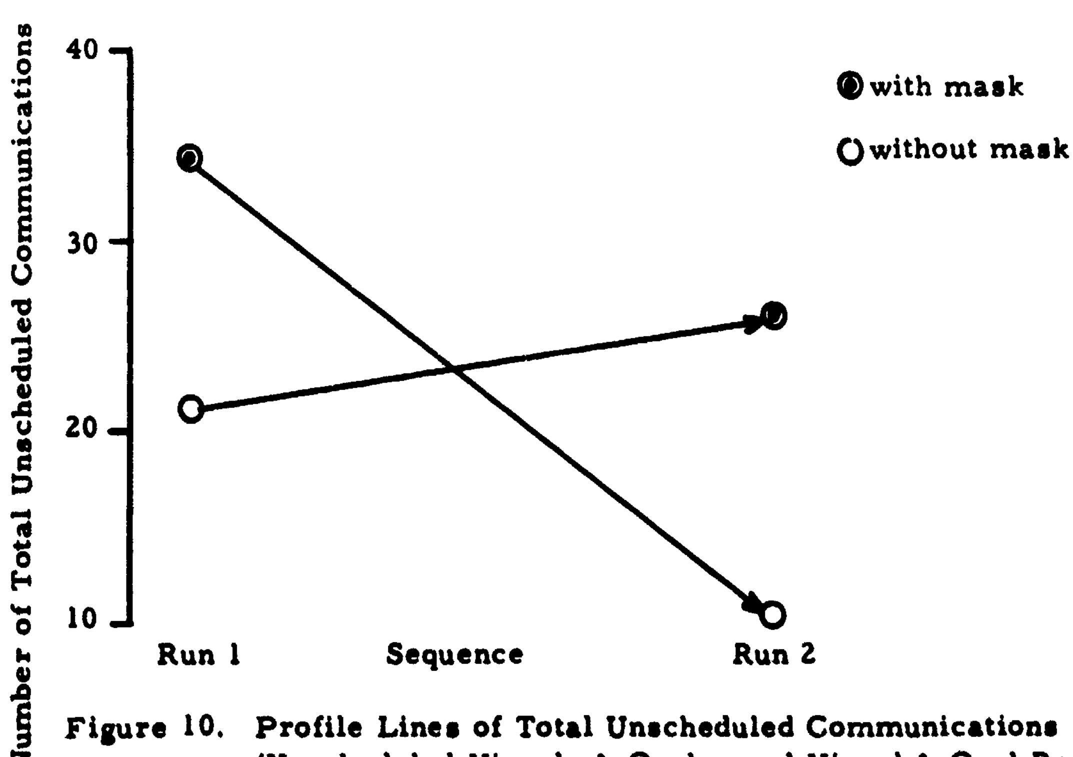


Figure 10. Profile Lines of Total Unscheduled Communications (Unscheduled Visuals & Orals, and Visual & Oral Repeats).

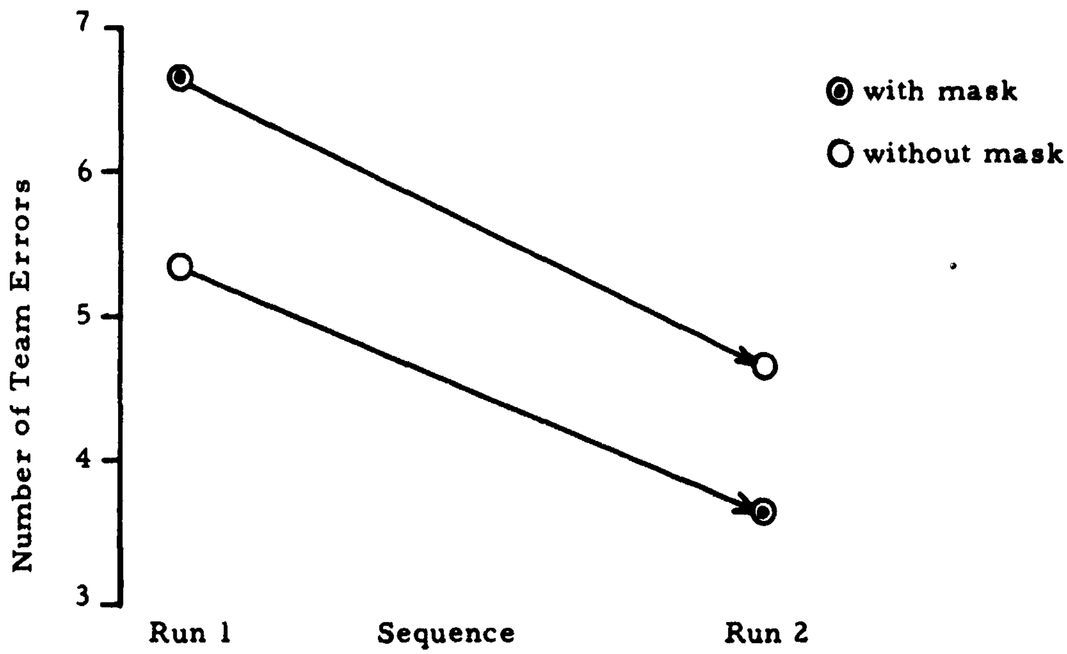


Figure 11. Profile Lines of Team Error per Task, by Sequence.

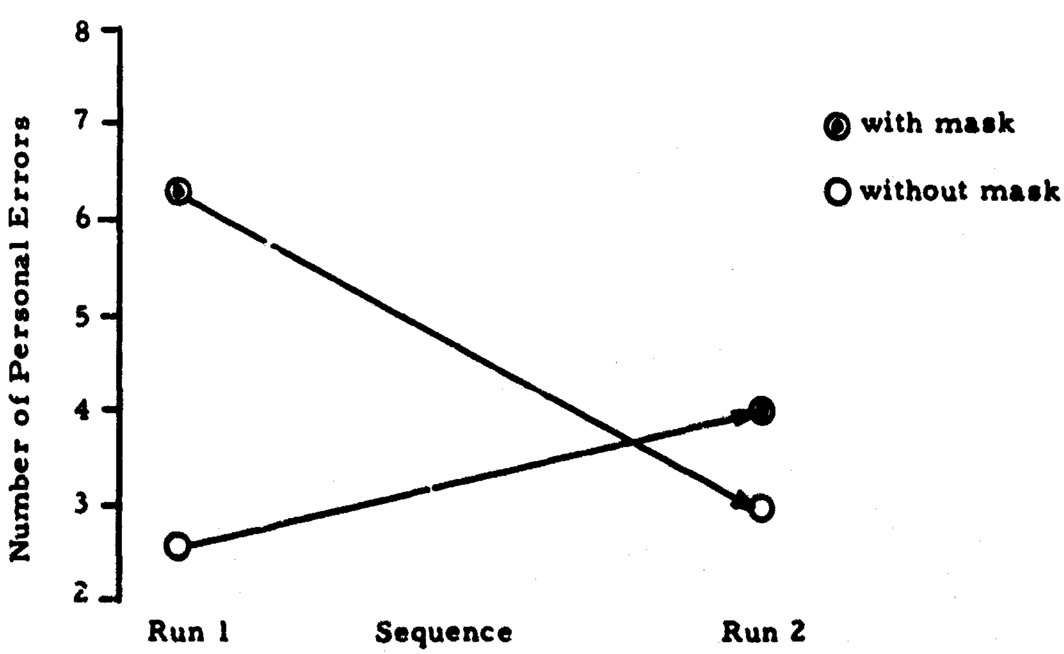


Figure 12. Profile Lines of Personal Errors per Task, Considering Sequence.

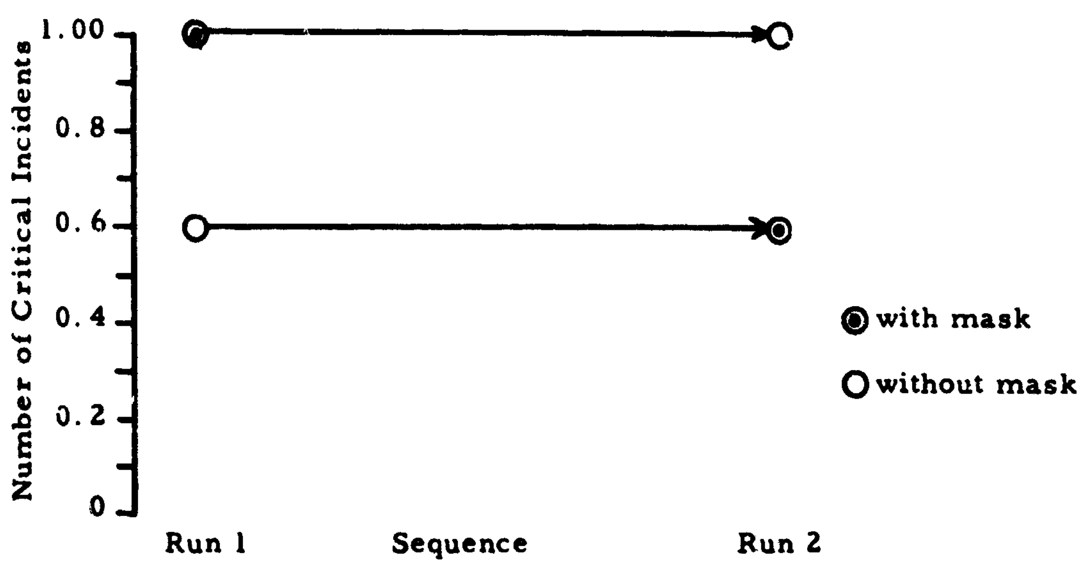


Figure 13. Profile Lines of Critical Incidents per Task.

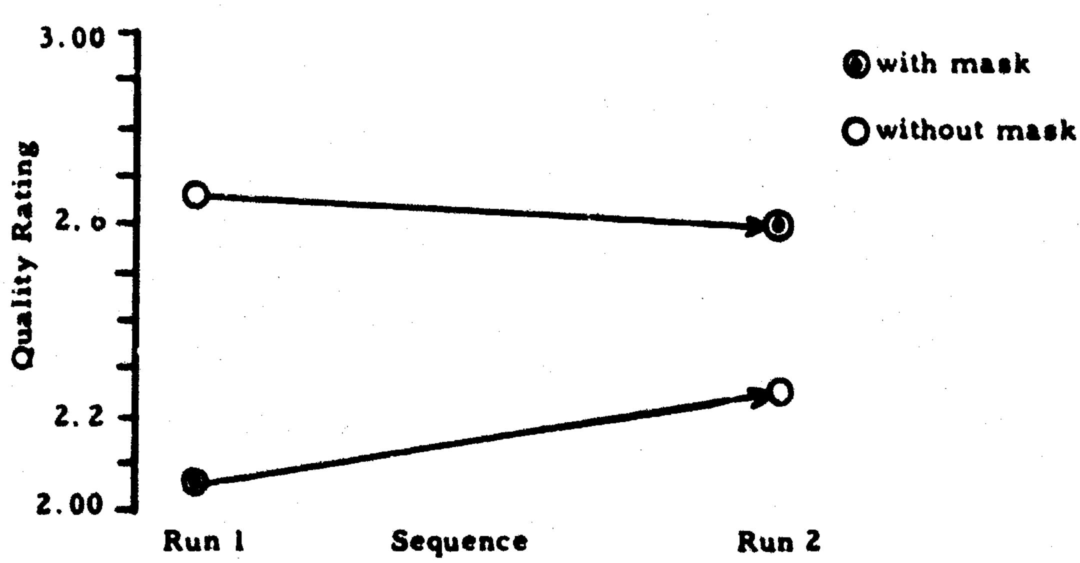


Figure 14. Profile Lines of Quality Rating per Task.

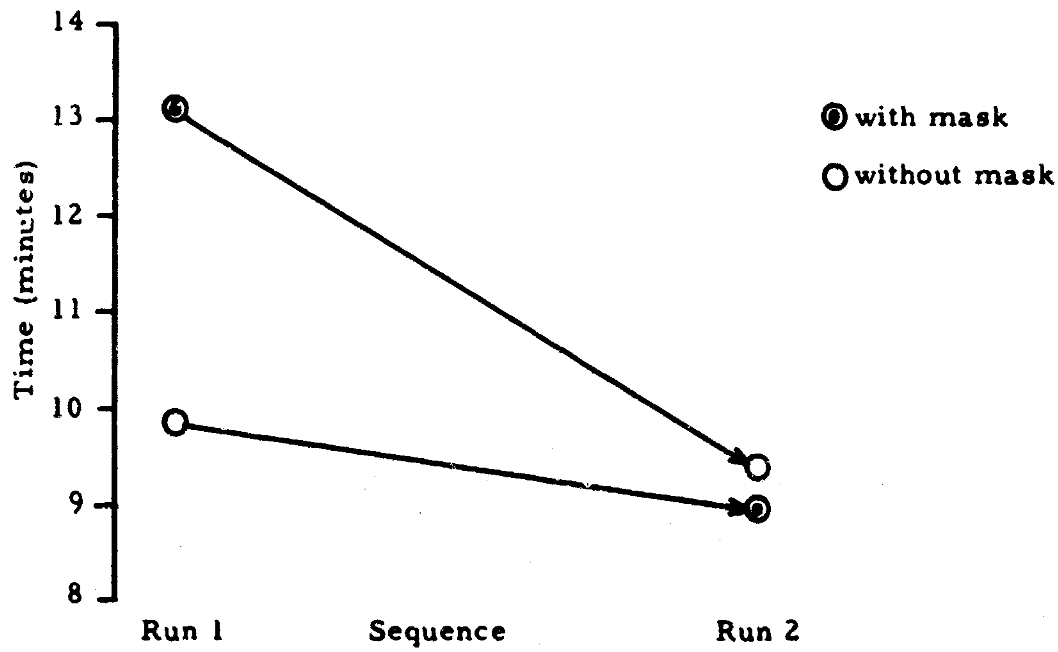


Figure 15. Profile Lines of Total Time per Task.

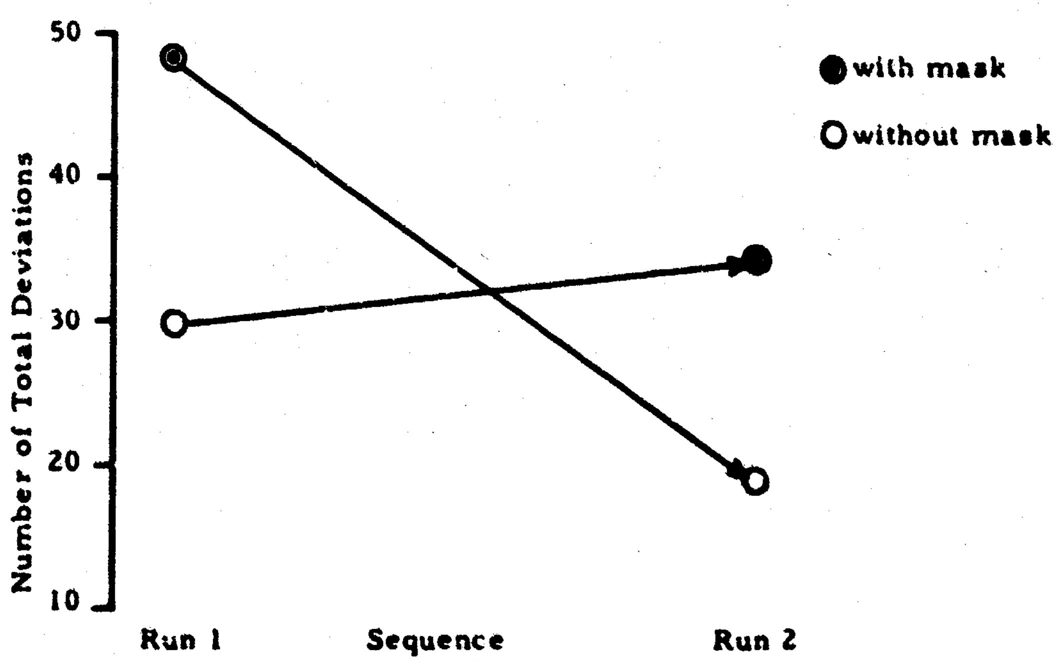


Figure 16. Profile Lines of Total Deviations (Total Error, Unscheduled Visual & Visual Repeat, and Unscheduled Oral & Oral Repeat).

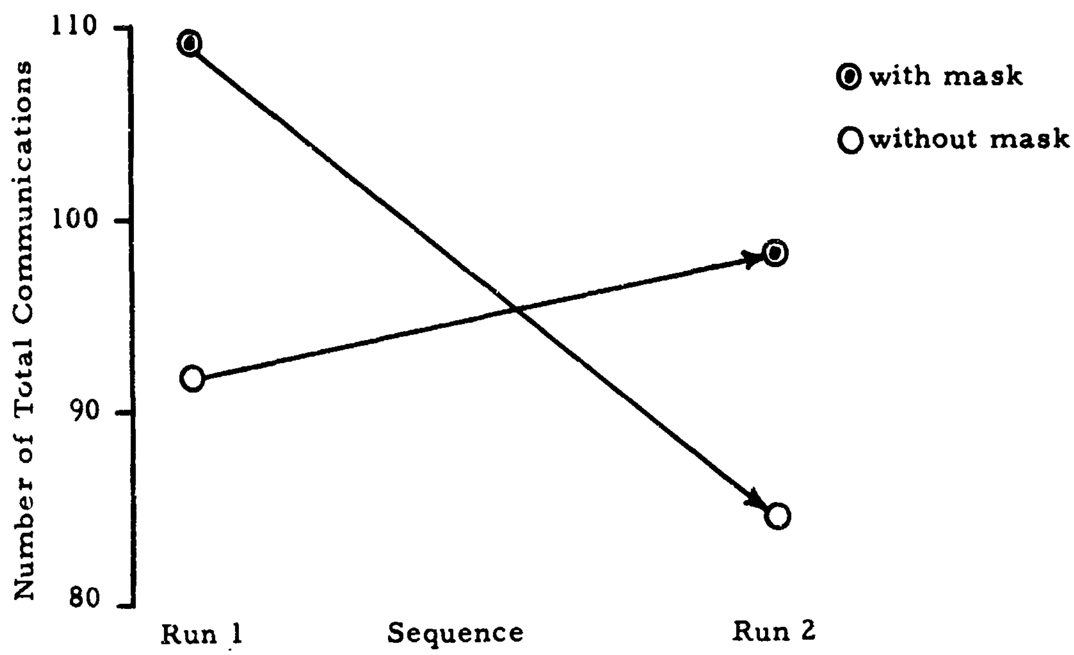


Figure 17. Profile Lines of Total Communications per Task (Total Visual and Total Oral)

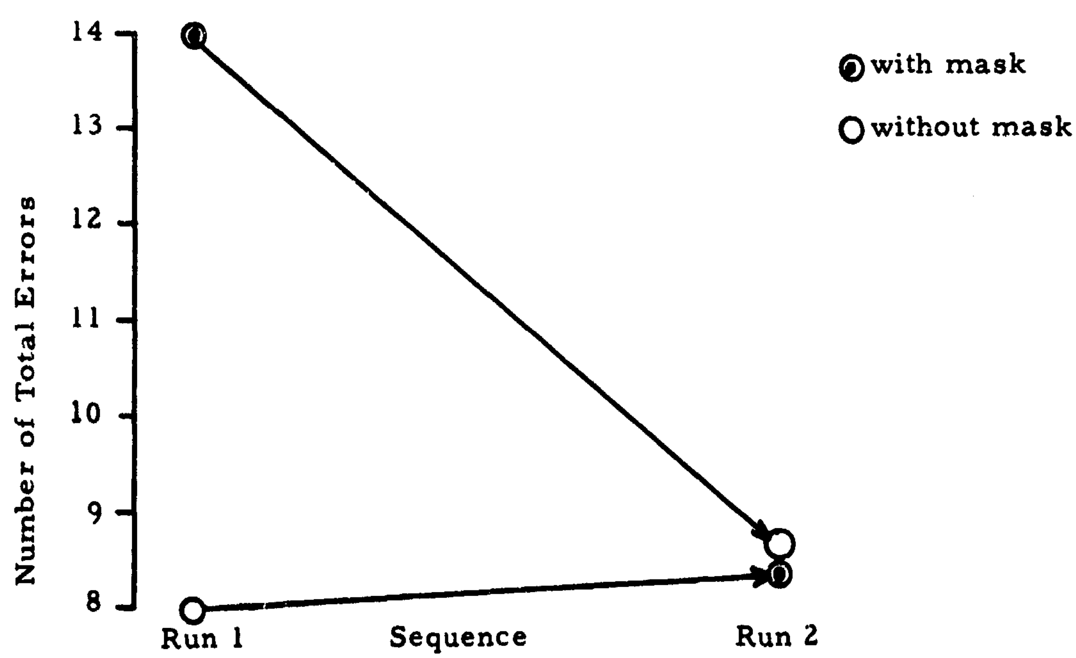


Figure 18. Profile Lines of Total Errors per Task (Personal Errors, Team Errors, and Unsafes).

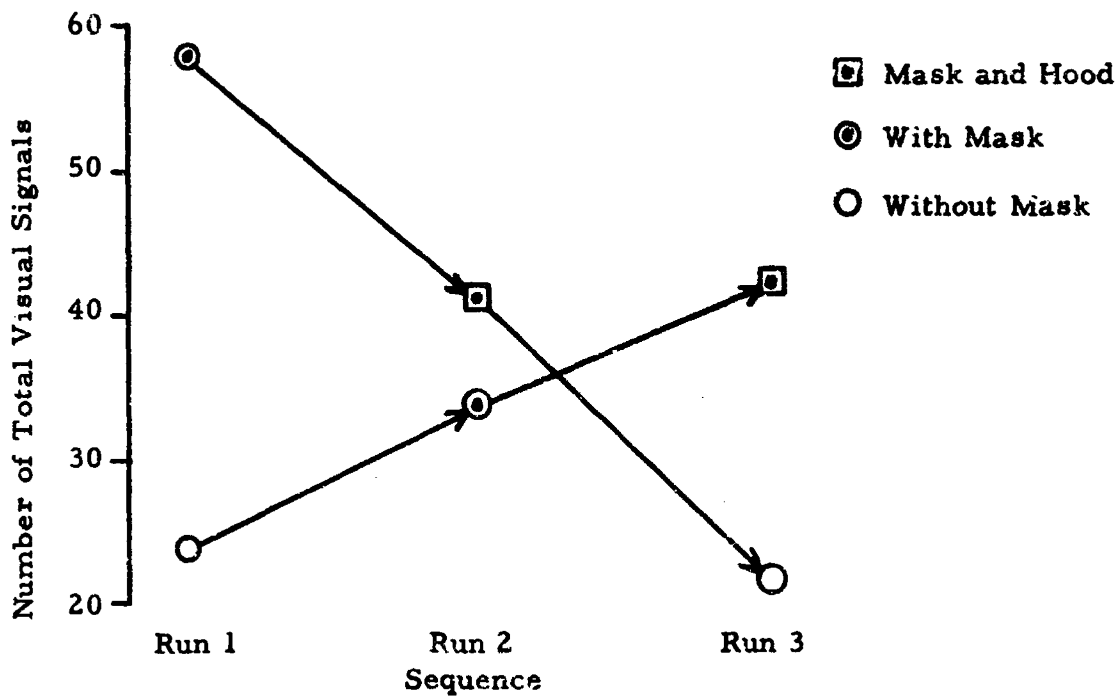


Figure 19. Profile Lines of Total Visual Signals per Task

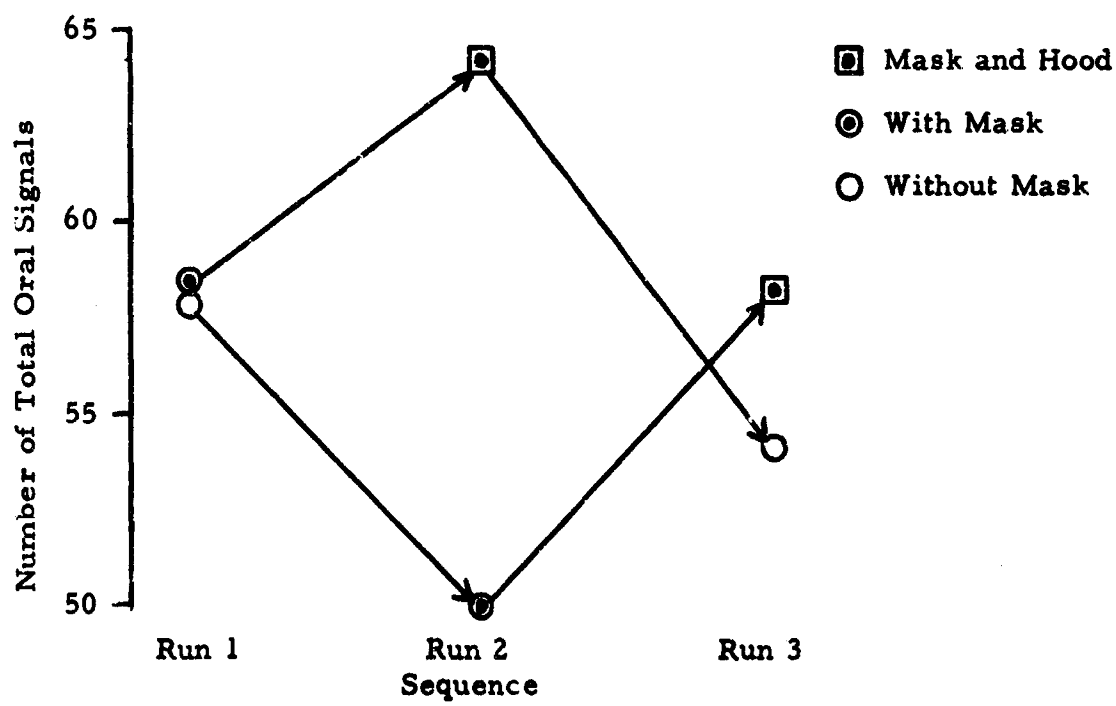


Figure 20. Profile Lines of Total Oral Signals per Task

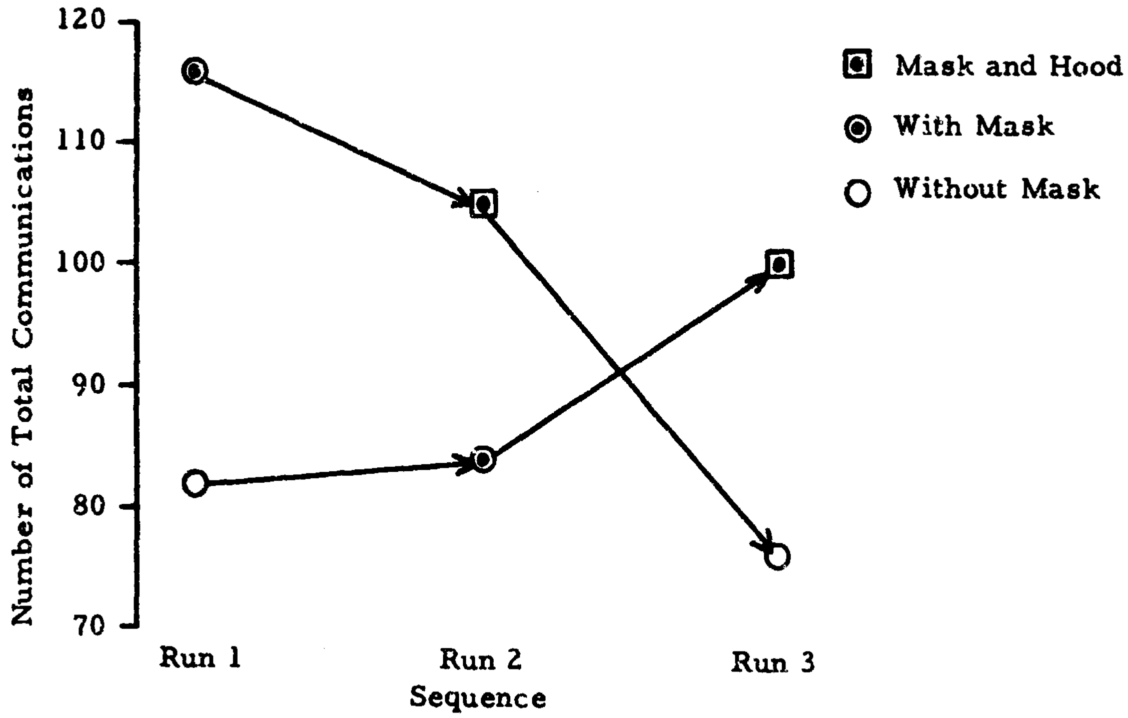


Figure 21. Profile Lines of Total Communications per Task

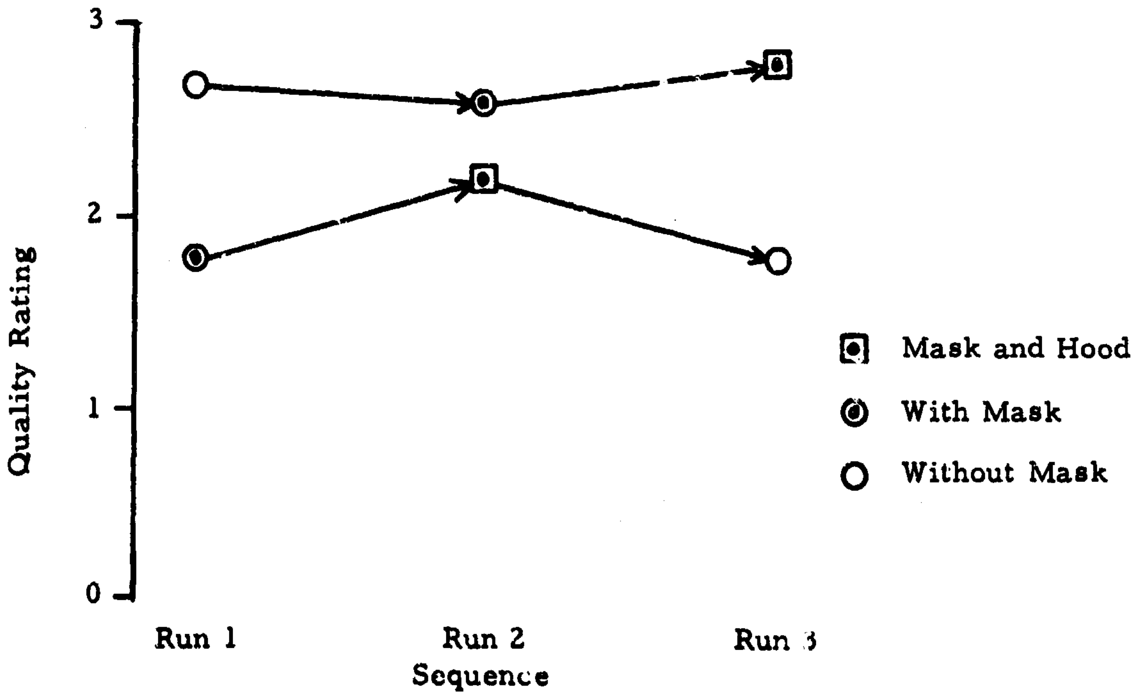


Figure 22. Profile Lines of Quality Rating per Task

APPENDIX C

SUBTASK DATA

The procedure for a standard reduction of data into totals by subtask was explained in Section IV. B. 2. a. The Subtask Bar Graphs on the following pages present each separate variable in its relative effect, in both conditions, during each of the five field subtasks described in Section II. B. 2.

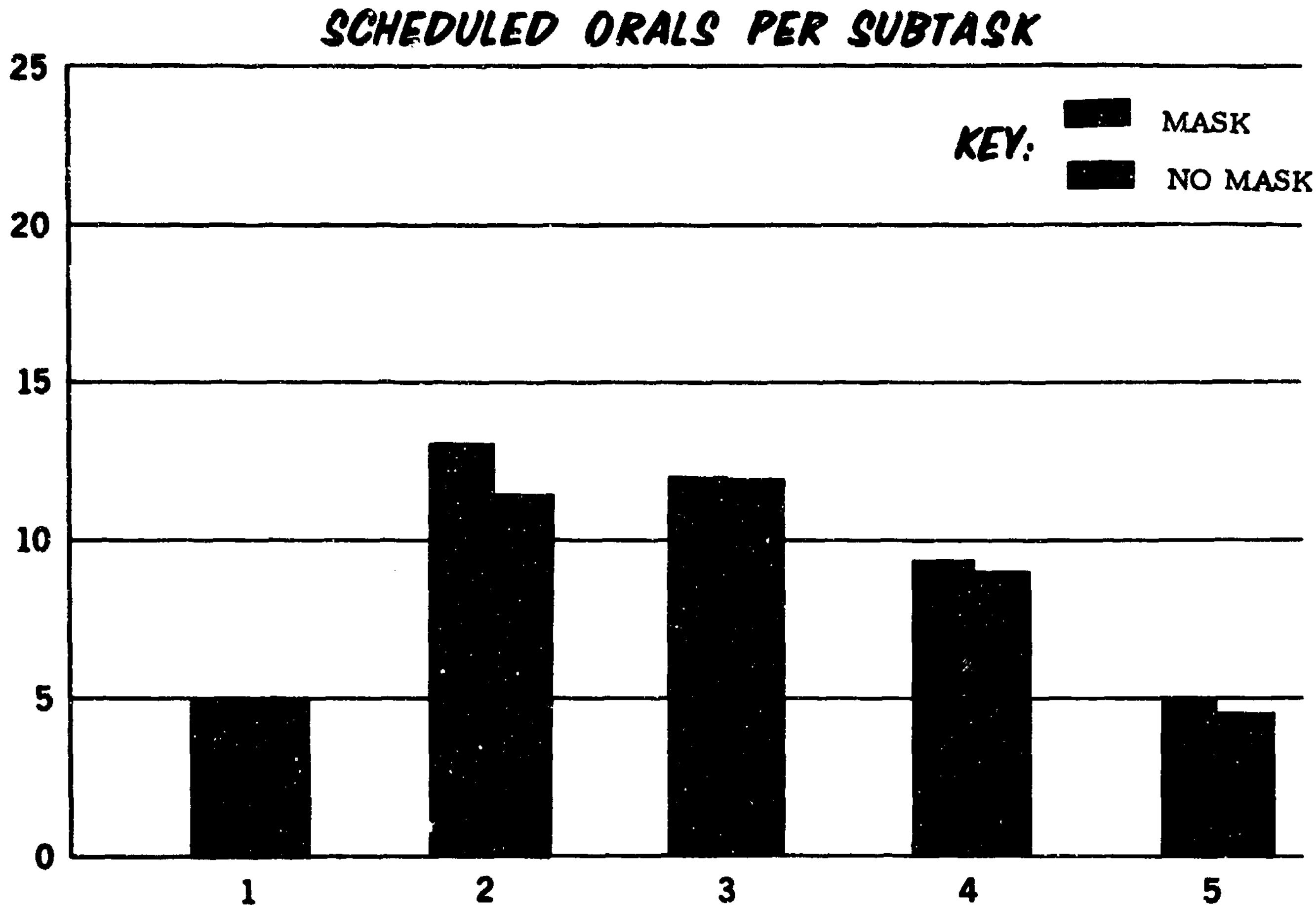


Figure 1. Average number of scheduled oral signals per subtask.

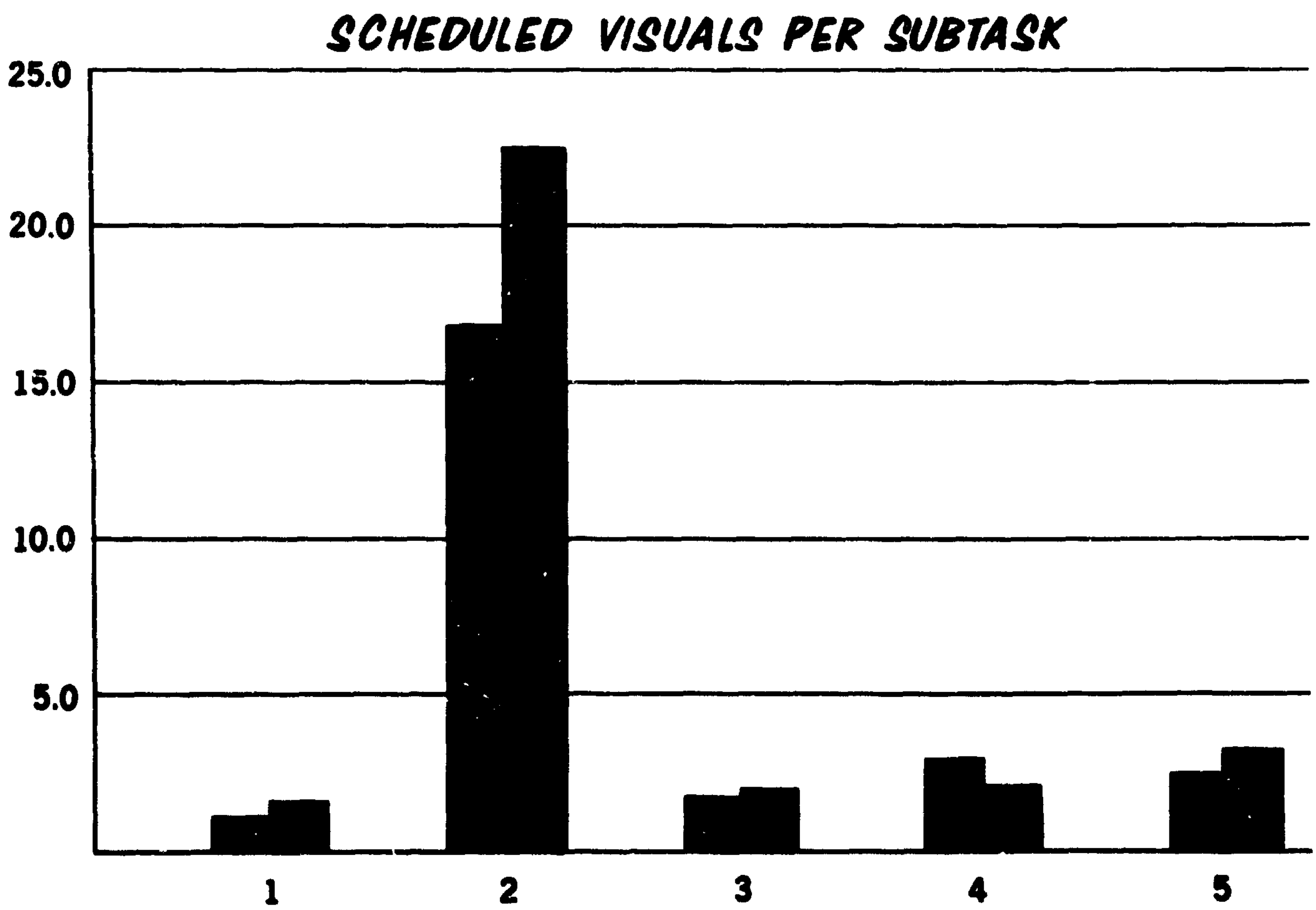


Figure 2. Average number of scheduled visual signals per subtask.

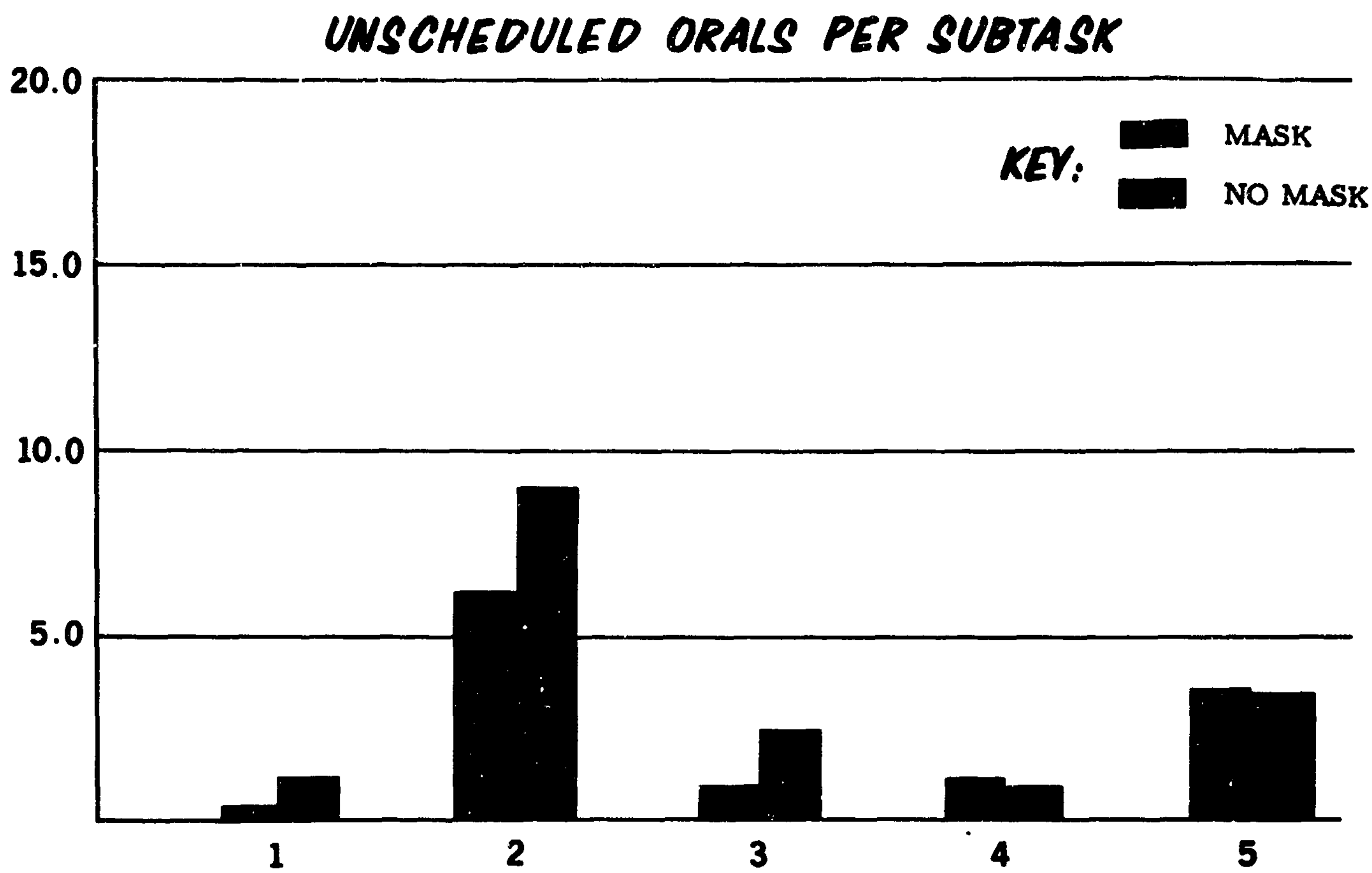


Figure 3. Average number of unscheduled oral signals per subtask.

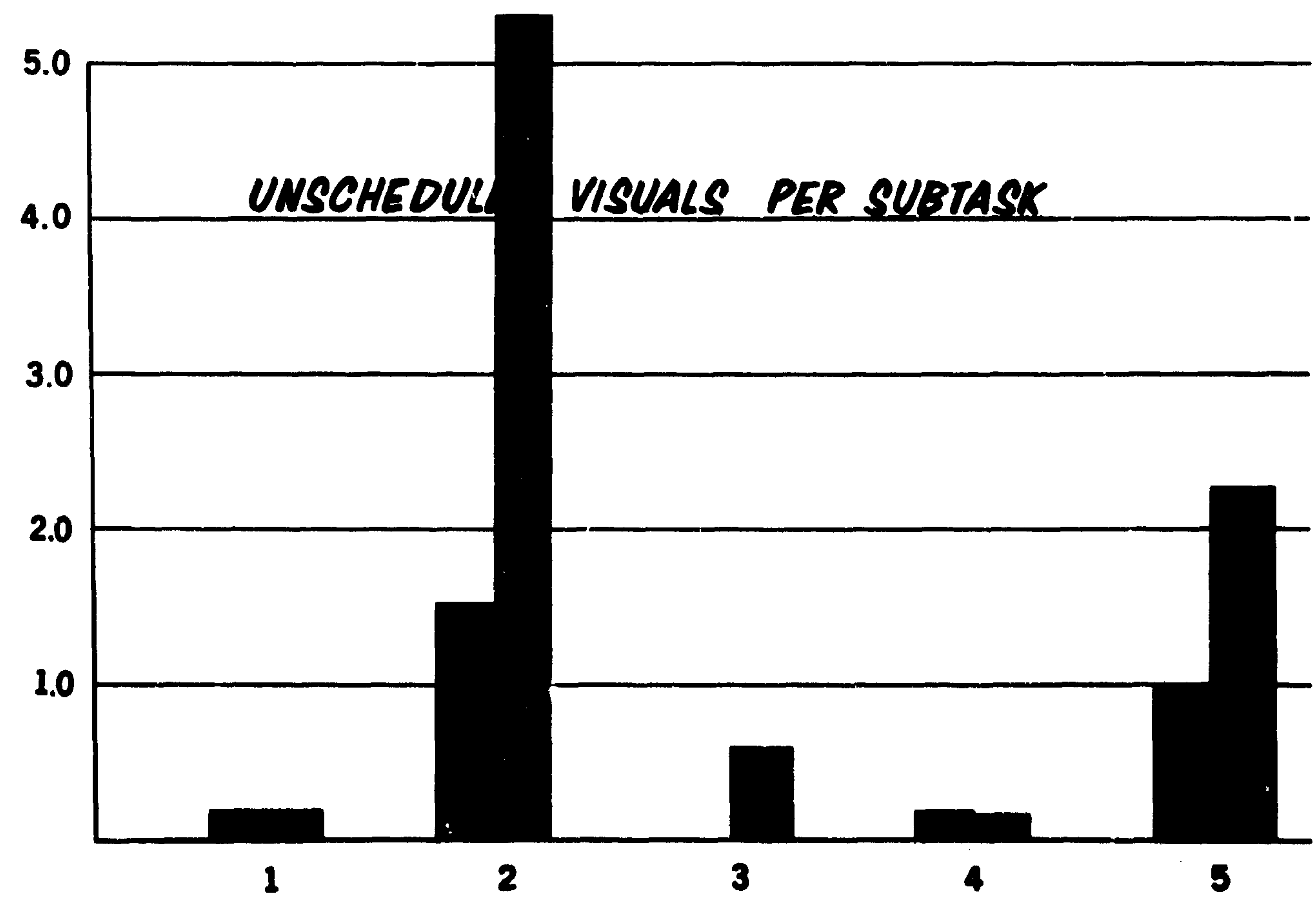


Figure 4. Average number of unscheduled visual signals per subtask.

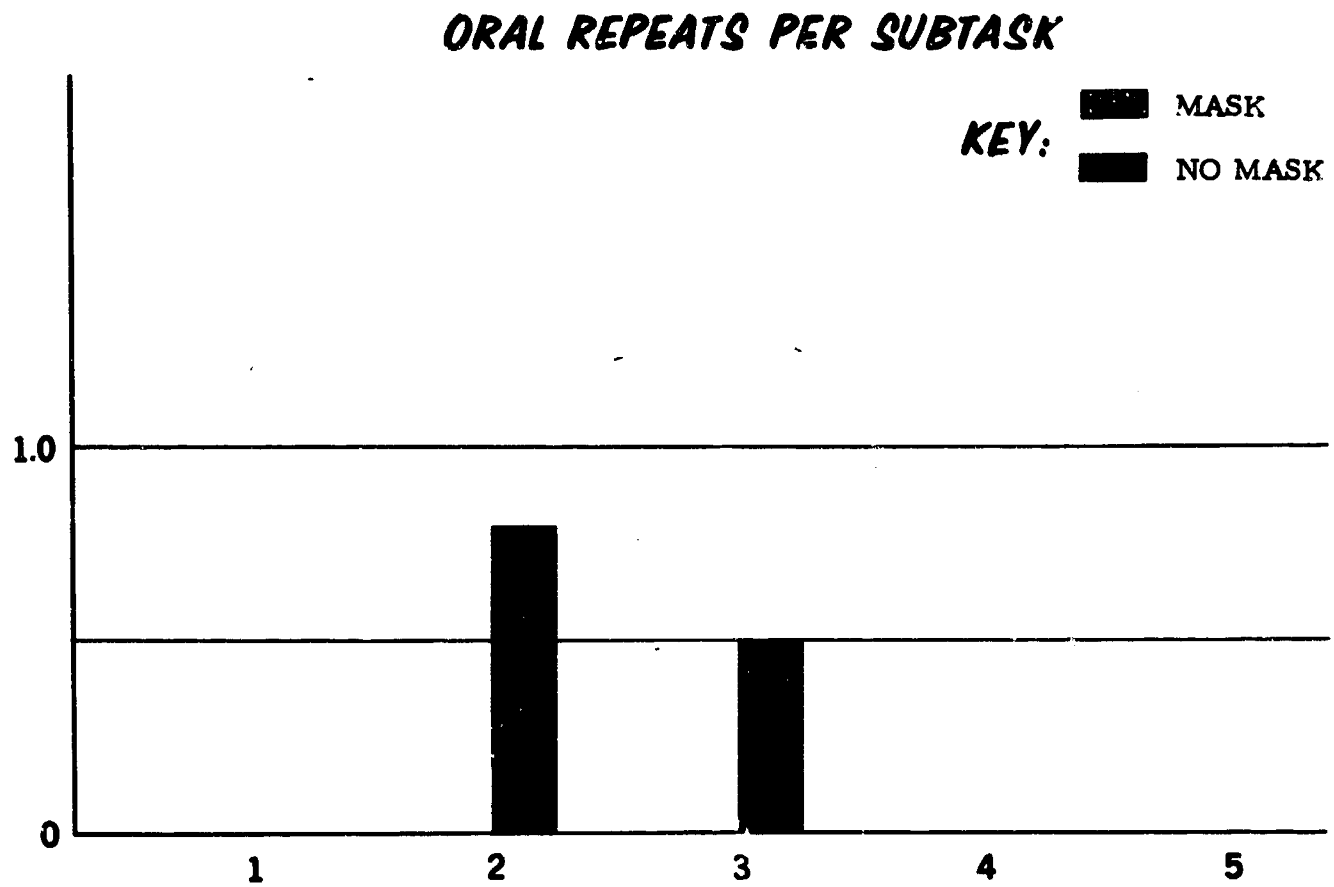


Figure 5. Average number of oral signals repeated per subtask.

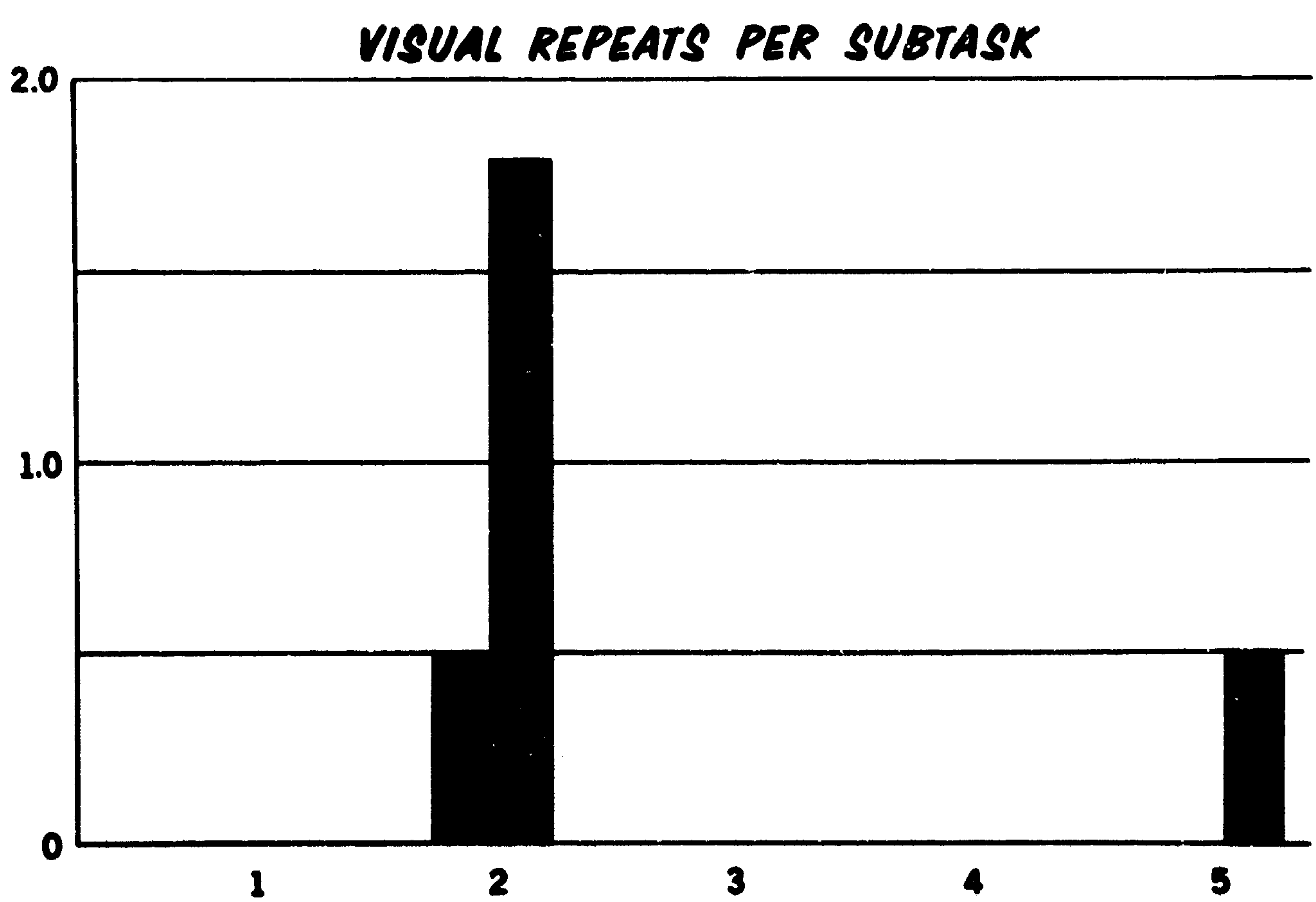


Figure 6. Average number of visual signals repeated per subtask.

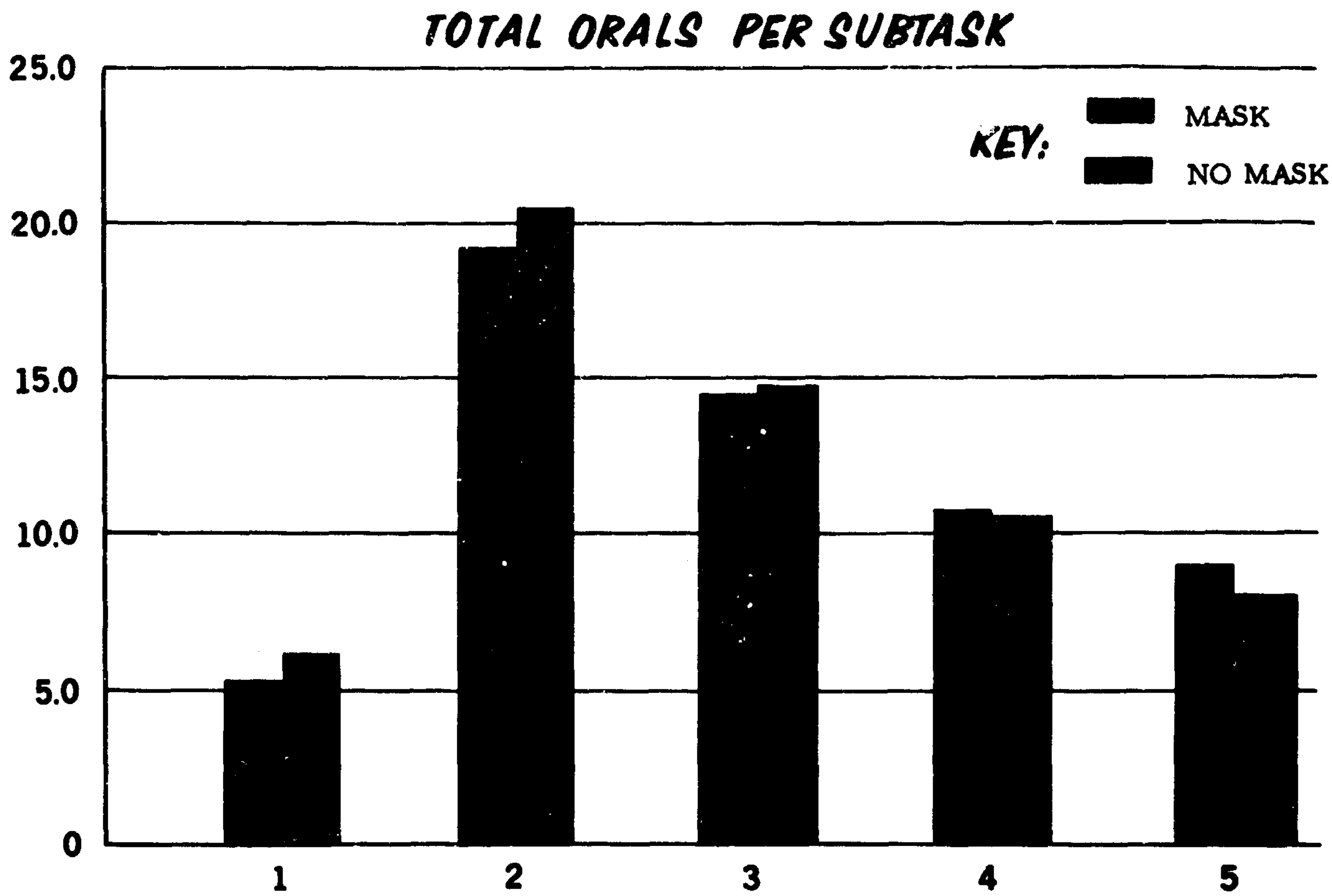


Figure 7. Average number of total oral signals per subtask.

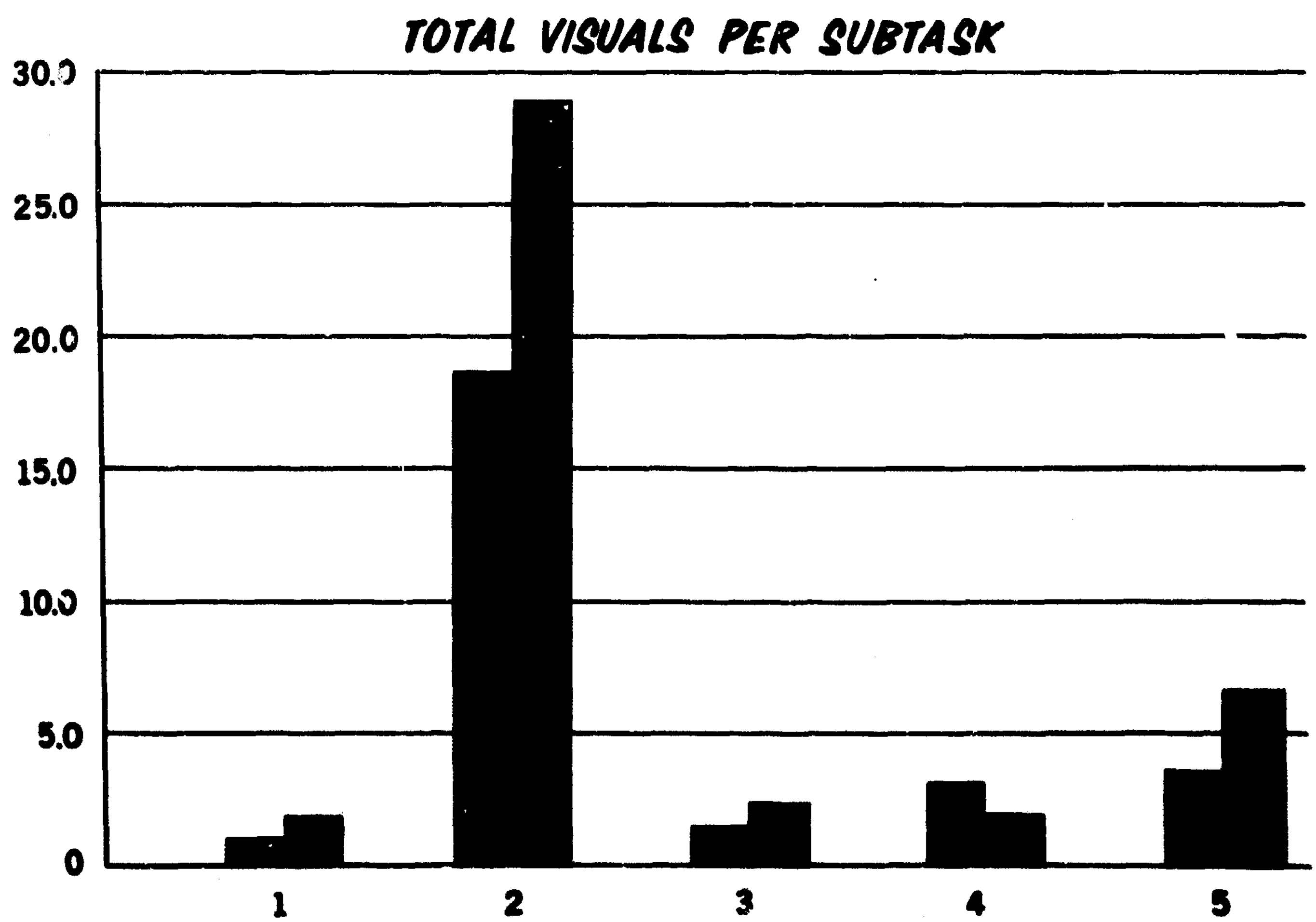


Figure 8. Average number of total visual signals per subtask.

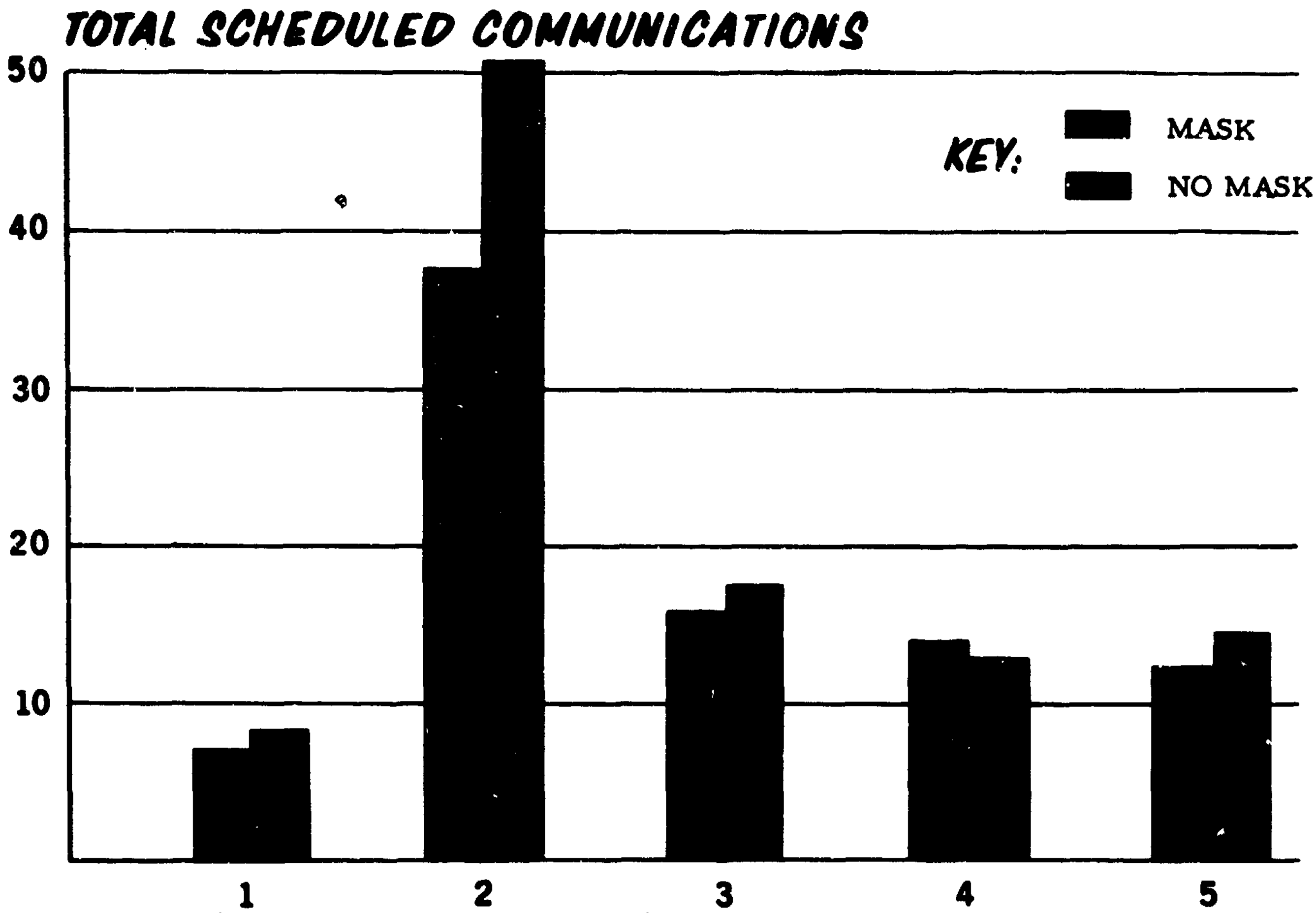


Figure 9. Average number of total scheduled signals per subtask.

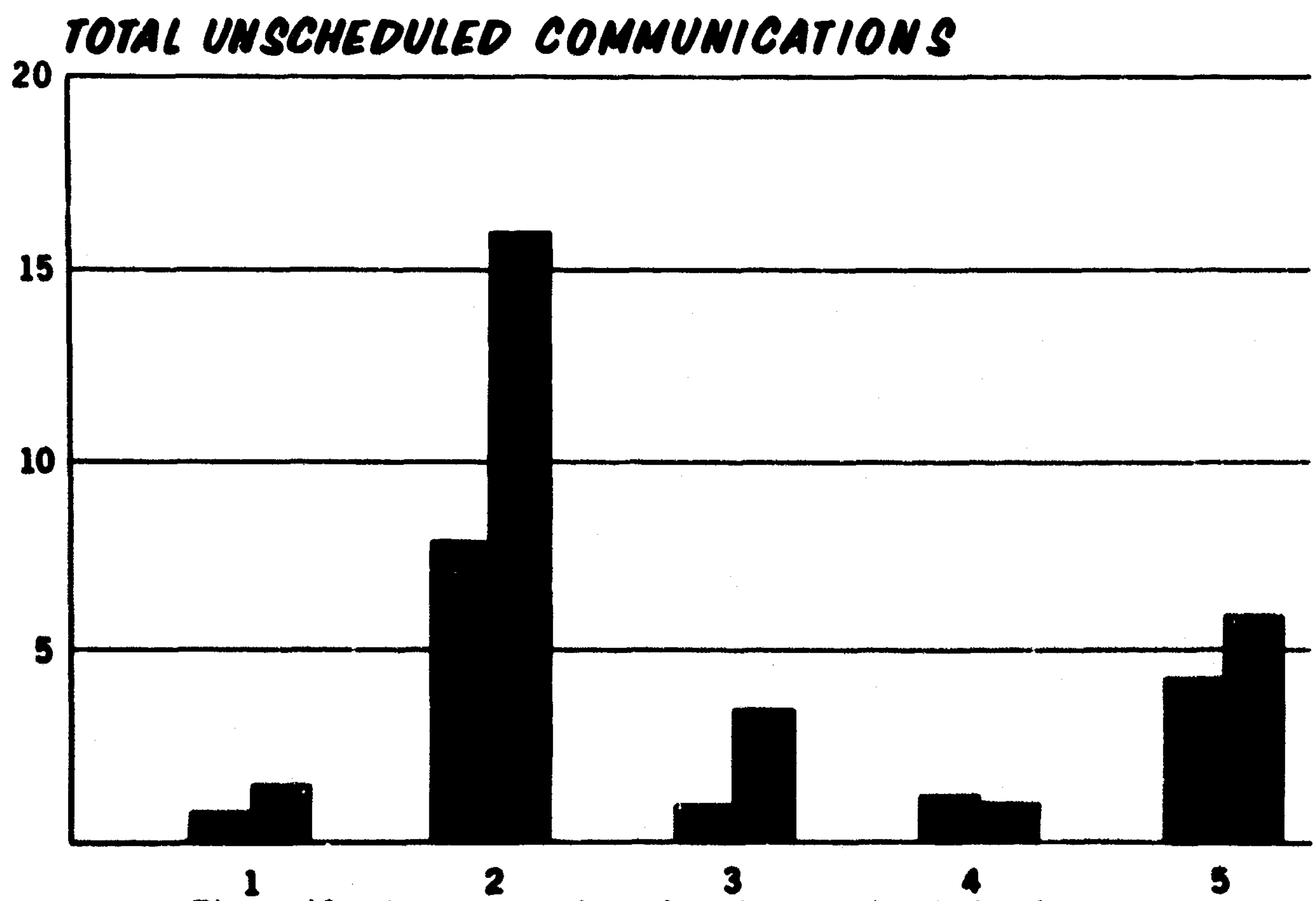


Figure 10. Average number of total unscheduled signals per subtask.

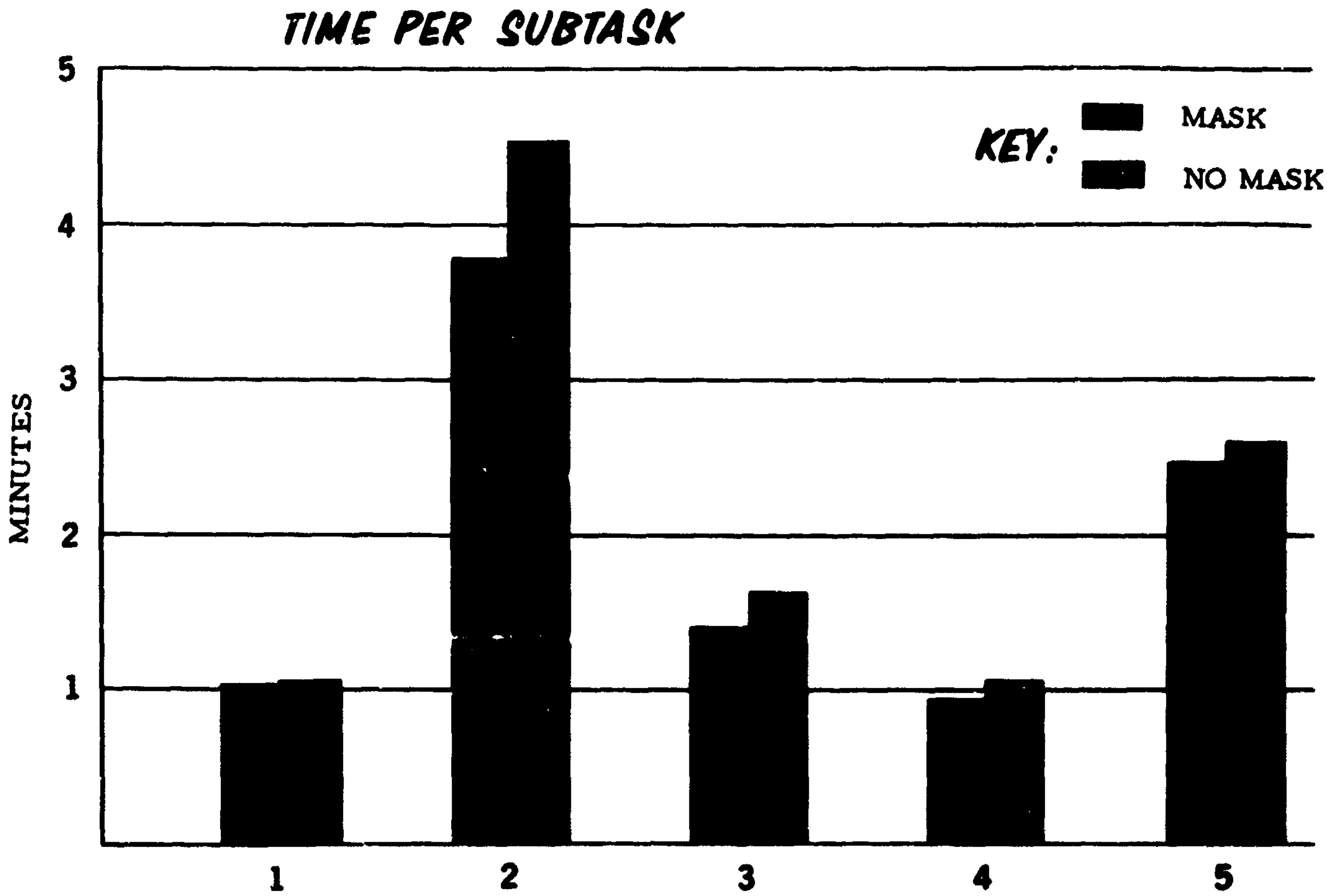


Figure 11. Average time per subtask.

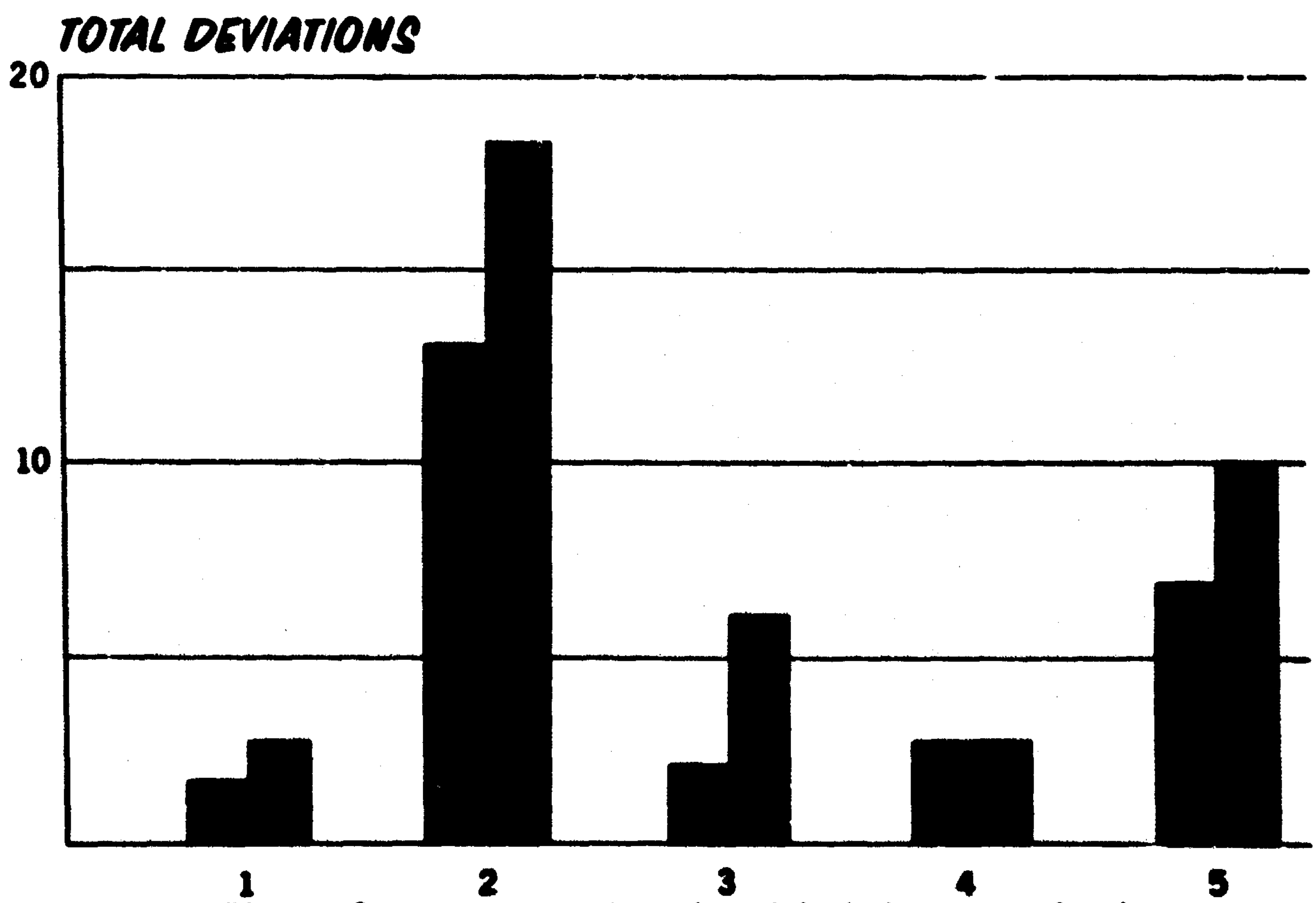


Figure 12. Average number of total deviations per subtask.

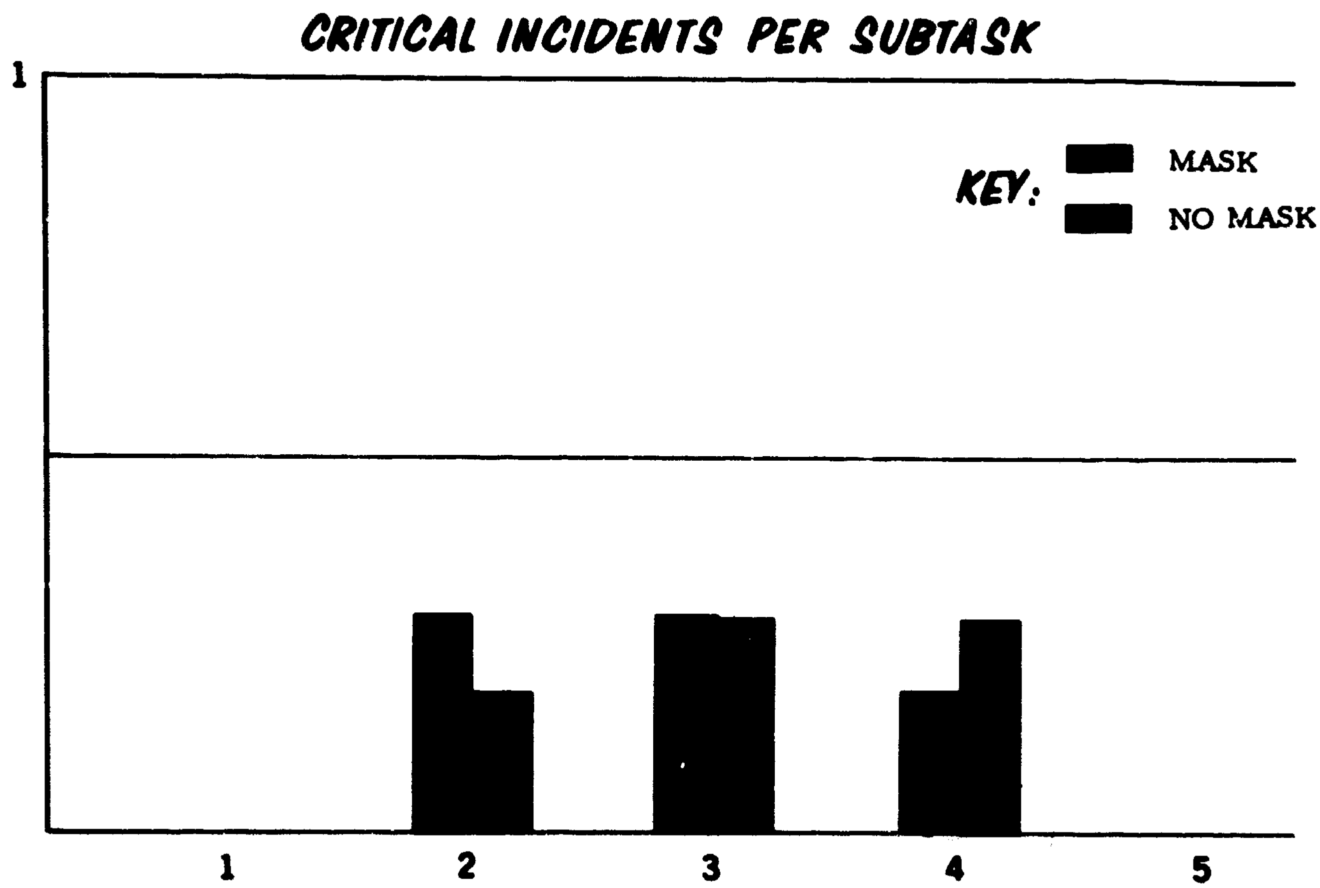


Figure 13. Average number of critical incidents per subtask.

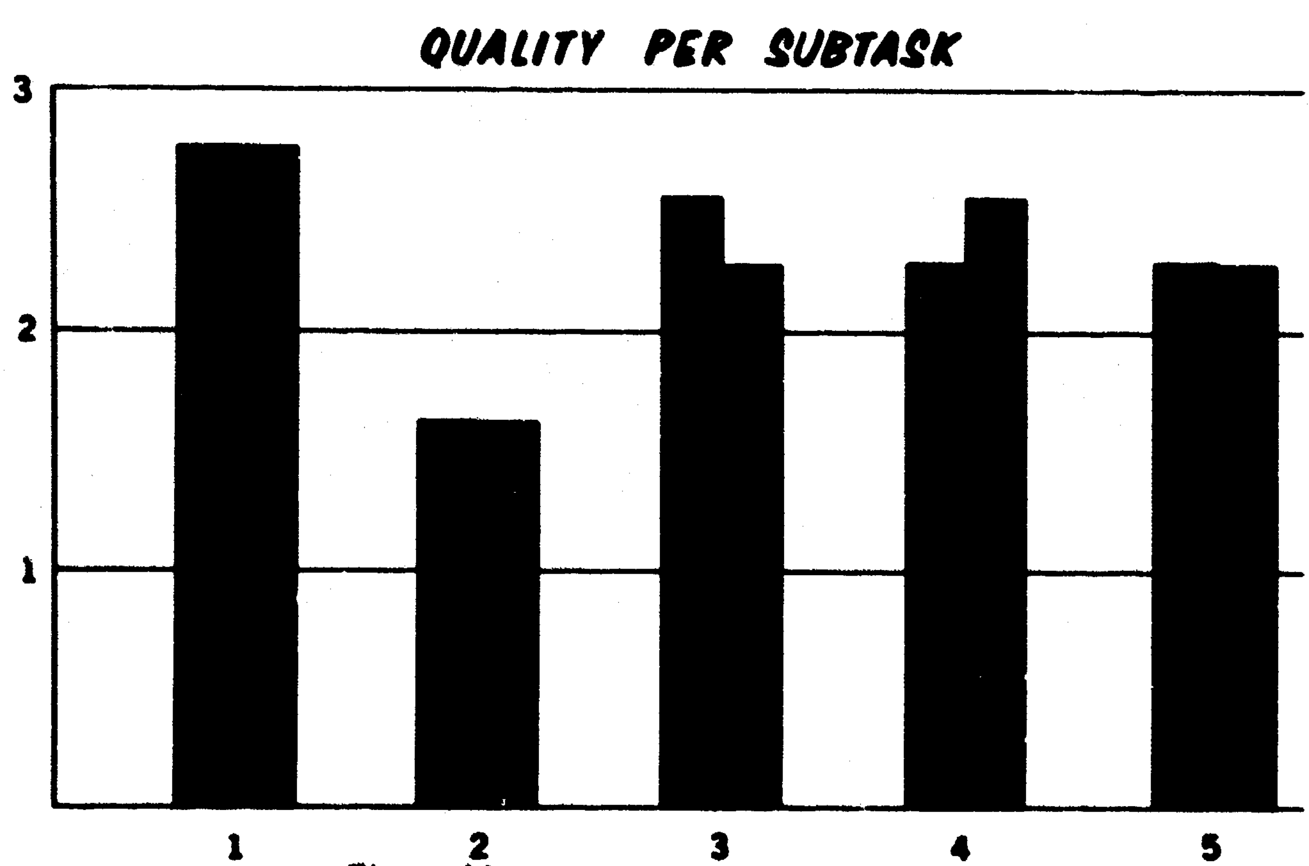


Figure 14. Average quality rating per subtask.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Dunlap and Associates, Inc. One Parkland Drive Darien, Connecticut 06821	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP N/A

3. REPORT TITLE

PERFORMANCE MEASSURES FOR HUMAN FACTORS ENGINEERING EVALUATION OF EARL EQUIPMENT

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Comprehensive Report SSD66-332 (618)

5. AUTHOR(S) (First name, middle initial, last name)

Dunlap and Associates, Inc.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
30 June 1966	139	80

8a. CONTRACT OR GRANT NO. DA-18-035-AMC-702(A) b. PROJECT NO. 1C622401A097 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) SSD66-332 (618)/FCR 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
Medical defensive aspects	Medical Research Laboratory, Research Laboratories, Edgewood Arsenal, Maryland 21010

13. ABSTRACT

(U) A simple measurement design is inadequate for field testing of teams. Observations must be structured to detect separate patterns of performance across multiple variables, individual team members, and various subtasks. Human observers at the field site cannot alone record this wide range of data at the pace of team performance. Data recording equipment, particularly videorecorders, assure a more adequate and objective data collection. Performance of certain team tasks, generally those requiring individual initiative in coordinating activity, may be significantly altered by the wearing of protective masks. Other tasks of a repetitive drill nature may be much less affected. Even very meticulous testing during field training activity may fail (because of the "slack time" and indirect motivations inherent in the practice environment) to reflect validly potential decrements in a tactical environment. For 105-mm howitzer crews specifically, there will probably be no consistently critical performance problems while wearing protective masks, although the team's normal pattern of oral intercommunication is definitely disrupted.

DD FORM 1473

UNCLASSIFIED

Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Human factors						
Performance measures						
Engineering evaluation						
EARL equipment						
Respiratory protective equipment						
Physiological effects						
EARL protective masks						
Test design						
Multiple regression analysis						
Intercorrelations						
Protective clothing						
Field tests						
Team tests						
Protective equipment						
Audio/visual recordings						
Intercommunications						
Heart rate						
Time						
Measurement techniques						
Videotaping techniques						

UNCLASSIFIED

Security Classification